DOCUMENT RESUME

ED 067 643 24 CS 000 179

AUTHOR	Puff, C. Richard
TITLE	Conceptual and Subjective Organizational Processes in Human Learning and Memory. Final Report.
INSTITUTION	Franklin and Marshall Coll., Lancaster, Pa.
SPONS AGENCY	National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
BUREAU NO	BR-7-0296
PUB DATE	Sep 72
GRANT	OEG-2-9-480269-1046
NOTE	99p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Cognitive Processes; Concept Formation; *Learning Processes; *Memory; *Retention; Verbal Stimuli

ABSTRACT

This report describes seven basic experiments designed to further elucidate the nature and function of two types of organization imposed by subjects in a free-recall type of memory task. The experiments involved verbal stimulus materials and employed college students as subjects. The two types of organization investigated were category clustering, or the tendency to remember conceptually related items together, and subjective organization, or the tendency for subjects to develop a relatively consistent sequence of responses across successive recall attempts. The main focus of several studies was the evaluation of the theoretical view that the amount that can be remembered is critically dependent upon the degree of organization imposed in recall by the subject. While three studies supported this view, the results of two experiments ran counter to it. These findings suggest that a general proposition that memory is crucially dependent upon such organization is, at best premature. Studies comparing the two types of organization indicated that conceptual clustering was used more extensively than was subjective organization. Furthermore, the two types of organization appeared to reflect relatively independent strategies, and similarities were found in their temporal characteristics. (Author)

ED 067643

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Final Report

Project No. 7-0296
Grant No. OEG-2-9-480269-1046

Conceptual and Subjective Organizational Processes in Human Learning and Memory

C. Richard Puff

Franklin and Marshall College

Lancaster, Pennsylvania

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
National Center for Educational Research and Development

FILMED FROM BEST AVAILABLE COPY

CS 000 179

Preface

The project covered in this final report comprised a series of seven experiments which were completed in stages throughout the project. Each of the experiments has been described in full detail in a separate interim report. A listing of these interim reports and other previous dissemination activities is provided in Appendix A along with some indication of the relationship between these reports and the originally proposed research. Some of the detail included in the interim reports is not repeated again since the major attempt here is to treat the project as a whole, placing the experiments in broader context, and emphasizing the integration of their findings.

Abstract

This report describes seven basic experiments designed to further elucidate the nature and function of two types of organization imposed by subjects in a free-recall type of memory task. The experiments involved verbal stimulus materials, and employed college students as subjects. The two types of organization investigated were category clustering, or the tendency to remember conceptually related items together, and subjective organization, or the tendency for subjects to develop a relatively consistent sequence of responses across successive recall attempts. The main focus of several studies was the evaluation of the theoretical view that the amount that can be remembered is critically dependent upon the degree of organization imposed in recall by the subject. While three studies supported this view, the results of two experiments ran counter to it. These findings suggest that a general proposition that memory is crucially dependent upon such organization is, at best, premature. Studies comparing the two types of organization indicated that conceptual clustering was used more extensively than was subjective organization. Furthermore, the two types of organization appeared to reflect relatively independent strategies, and similarities were found in their temporal characteristics. A number of types of scores for expressing organization were also investigated.

Table of Contents

Page

Abstract i

Preface v

List of Tables ix

Introduction 1-1

Experiment 1: The Role of Clustering in Free Recall 2-1

Experiment 2: An Investigation of Two Forms of Organization in Free Recall 3-1

Experiment 3: Free Reclll With Serial and Simultaneous Presentation of Categorized Word-Lists . . . 4-1

Experiment 4: An Investigation of Memory Performance With Three Types of Presentation Sequences and Seven Measures of Organization 5-1

Experiment 5: Temporal Properties of Organization in Recall of Unrelated Words 6-1

Experiment 6: Effects of Instructions Upon Recall and Clustering With Conceptually Categorized Materials 7-1

Experiment 7: Effects of Categorical and Non-Categorical Context Upon Recall of the Same Stimulus Words 8-1

General Discussion and Conclusions 9-1

References 10-1

Appendix A: Listing of Experiments Performed, Relationship to Original Proposal, and Dissemination Activities 11-1

Appendix B: Brief Formulas for the Clustering Measures . .11-3

List of Tables

Table		Page
1.	Mean Number of Items Recalled from the Practice List, the Non-Categorized List, and the Categorized List in Experiment 1	2-3
2.	Group Means of Observed Clustering and Intertrial Organization Expressed as Percentages of Their Maximum Possible Amounts in Experiment 2	3-3
3.	Group Means of Observed Clustering and Intertrial Organization Expressed as Deviations from Chance Expected Values in Experiment 2	3-5
4.	Group Means of Observed Clustering and Inter trial Organization Expressed as Percentages of Their Maximum Possible Amounts for High and Low Strength Category Lists in the Seond Study of Experiment 2	3-7
5.	Mean Recall Measures Obtained with Total Exposure Time Equated between Serial and Simultaneous Methods of Presentation in the First Study of Experiment 3	4-4
6.	Mean Recall Measures Obtained with Total Processing Time Equated between Serial and Simultaneous Methods of Presentation in the First Study of Experiment 3	4-6
7.	Proximity Measures and Sequential Codes for the Three Types of Sequences Used in Experiment 4	5-5
8.	Group Means of Number of Items Correctly Recalled in Experiment 4	5-6
9.	Group Means of Various Clustering Measures and Results of Analysis of Variance for Each Measure in Experiment 4	5-8
10.	Group Means of Subjects Median Interword Times (Without Regard to Organization) for Five Stages of the Recall Sequence Within Trials, and for Three Blocks of Trials, in Experiment 5	6-4
11.	Mean Density of Observed ITR Units as a Percentage of Opportunities. Results are for Five Stages of the Recall Sequence and for Three Blocks of Trials in Experiment 5	6-5

Table Page

12. Means of Subjects Median Interresponse Times for ITR and Non-ITR Units at Five Stages of the Recall Sequence and for Three Trial Blocks in Experiment 5 6-7

13. Results of Analyses of Interresponse Times Across Fifths of Output Within Each Trial Block in Experiment 5 6-8

14. Means of Subjects Median Interresponse Times for ITR and Non-ITR Units for Five Stages of the Recall Sequence and Three Blocks of Trials in Experiment 5 6-10

15. Mean Interresponse Times in Seconds Between Words in Units Which Were Repeated at Least X Times in Experiment 5 6-11

16. Mean Number of Words Recalled on Each Trial in Experiment 6 7-6

17. Mean Number of Categories Represented in Recall of the Ten Category List Under All Instructions in Experiment 6 7-7

18. Mean O-E(SCR) Clustering Scores for Each Trial in Experiment 6 7-9

19. Mean ARC Clustering Scores for Each Trial in Experiment 6 7-11

20. Mean Number of Clusters Observed in Recall of the Ten Category List Under All Conditions in Experiment 6 7-12

21. Some Characteristics of the Stimulus Lists and the Number of Subjects Presented With Each List in Experiment 7 8-4

22. Mean Recall Scores of Various Kinds of Items From Categorized and Non-Categorized Lists in Experiment 7 8-6

Introduction

The unifying objective of the research comprising this project was to further our understanding of the nature and function of organizational processes in human free-recall verbal-learning and memory.

In the basic laboratory paradigm of free recall the subject is presented with a list of stimulus items, or other body of material, which he is asked to try to remember, but he is free to recall the material in any order that is easy, or "natural", for him. This paradigm is becoming increasingly emphasized in both research and theory because of the opportunity that the subject has to restructure the material, or to impose organization upon it which is consistent with his own conceptual categories and previous experiences, etc. Thus, it can be argued that this kind of laboratory task is in at least some ways the most analogous to the typical kind of non-laboratory learning situation.

The major interest in the kinds of organizing processes reflected in free recall performance stems primarily from the theoretical view that they are critically involved in determining the ease of learning and remembering a body of stimulus material (e.g., see Mandler, 1967, 1968; and Tulving, 1962a, 1968). Still further evidence of the growing significance of this theoretical view is shown by, for example, Spitz's (1966) proposition that the major functional deficiency of mentally retarded children is in the area of such organizing processes, and by the similar position taken by Jensen (1971) that lower levels of ability for such processes represent the crucial cognitive deficit of culturally deprived children. Thus it is becoming quite generally accepted that organization facilitates learning and retention.

The empirical evidence in support of this view is, however, far from being totally compelling. For example, much of the basic support claimed for this position is really correlational in nature. It is typically observed that both recall and organization increase with successive practice trials (e.g., see Bousfield, Puff & Cowan, 1964) and that subjects who organize to a greater degree also recall more material (e.g., Tulving, 1962b). These correlational data, of course, do not really permit the inference of a cause and effect relationship. Furthermore, there is other evidence which pretty straightforwardly contradicts the view that recall is dependent upon organization (e.g., Cofer, 1967). Particularly striking evidence against this view occurs when organization and recall vary inversely across experimental conditions (e.g., Allen, 1968).

The possibility of having identified a major determinant of human learning, and especially one which might delimit the problem involved in several populations of "slow learners", certainly justifies further basic research on organizational processes. Furthermore, the existence of an appreciable amount of contradictory evidence makes continued investigation all the more imperative.

The explanation of how organization might function to facilitate learning and memory is also not yet clearly established. The general theoretical orientation for much of the thinking on this topic has been Miller's (1956a, 1956b) unitization hypothesis. Briefly, Miller theorized that the human has an immediate memory capacity which is limited to 7±2 units. The specification of what constitutes a unit has been a hindrance to objective applications of this position, but it is assumed that in the situation where highly meaningful (i.e., very familiar) words are used as stimuli each unit initially consists of a single word. As a result of repeated practice with the same words, it is thought that several words come to be included within a single unit, called a higher-order unit, or "chunk", with the net result that the subject can then recall more than 7±2 words. This process of enriching the content of the units is assumed to continue until the entire stimulus list can be recalled.

Thus, organization of the material by the subject is thought to facilitate this process of unitizing, or chunking, by drawing together individual words on the basis of some conceptual relationship, meaningful similarity, or even an idiosyncratic association. It is most commonly further assumed that the words comprising a unit get stored together in the subject's memory (e.g., Tulving & Pearlstone, 1966). The organization seen at the time of recall then presumably reflects the subject's previous organizing activities and the consequent fact that the components of the same unit are stored together in memory. A somewhat different view of organization is held by other investigators like Bower, Clark, Lesgold, and Winzenz (1969) and Slamecka (1969). Briefly, in this view, organizing activities do not lead to the members of a unit actually being stored in memory together, but to the formation of a retrieval plan. The retrieval plan includes the information about which words constitute a unit, the hierarchical arrangement of units, etc. The organization actually seen in recall is thus thought to be a reflection of the use of the plan to guide the retrieval of the items from memory. In other words, the members of a unit should appear in close contiguity in recall because the plan directs the retrieval processes to them in close succession, rather than because they have actually been stored together in memory.

Further evidence bearing upon the distinction between these theoretical views is needed, not only just to achieve a better understanding of the basic mechanisms involved in organizational processes, but also because the two views suggest somewhat different techniques and conditions for maximizing organization, etc. In other words, since attempts to explore more applied uses of organization will be guided by theory as well as empirical findings, it is important to determine the theoretical base from which it is likely to be most fruitful to work.

Another of the substantive problems in the study of organizational processes concerns how the organization manifest in the subjects' recall protocols is most appropriately specified. There is no problem with the basic defining operations for the two types of organization which have been most widely studied.

The historically first, and most thoroughly investigated, form of organization is that called category clustering. Following the work by Bousfield (1953) this form of organization refers to the tendency for subjects to recall words from the same taxonomic, or conceptual, category together in runs, or clusters. More specifically, in the typical investigation of clustering the stimulus list might comprise, for example, ten members of each of the taxonomic categories of animals, vegetables, and pieces of furniture. The words from all categories would then usually be randomized into a single "scrambled" sequence and would be presented to the subjects one at a time. The subjects are instructed to try to learn the words, but that they are free to recall them in any order that seems easy or "natural" for them. Clustering is then observed when the subjects do not recall the words in the randomized order of presentation, but rather, tend to group together the members of the same category. The standard unit of measurement of clustering is the stimulus category repetition (SCR) and an observed unit of SCR, O(SCR), is scored each time a word from any category is directly followed in recall by another word from the same category.

The second major form of organization investigated to date is that called subjective organization by Tulving (1962a). It is also variously known as intertrial organization, sequential constancy, or seriation. The typical defining operations for the study of subjective organization begin with a list of "unrelated" words, i.e., words that are not categorically or associatively related according to the appropriate sets of norms. Multiple randomized orders of the list are prepared for presentation to subjects who have been given instructions for free recall. Then the words are presented one at a time according to the first randomization, and their recall is obtained from the subjects; the second randomization is presented and a second recall is obtained; and so on. Subjective, or intertrial, organization is shown when the subjects tend to develop a fixed order of recall across trials. Thus, the series of recall protocols for a single subject is scored in successive pairs (e.g., 1 & 2, 2 & 3, 3 & 4, etc.) to determine the amount of obtained intertrial repetition, O(ITR), as originally defined by Bousfield, Puff, and Cowan (1964). A unit of O(ITR) is traditionally scored each time two words recalled in direct succession on one trial are also recalled contiguously and in the same order on the next trial.

Thus, the two forms of organization are both soundly operationally defined and are distinctive from each other. Clustering is defined by the order of recall on a single trial and reflects the use of broad conceptual categories explicitly built into the stimulus list by the investigator. Subjective organization requires the comparison of the order of recall on two trials to determine the extent to which the order has been duplicated. Scoring for subjective organization is not based upon the use of categories built into the list by the experimenter, and may reflect units of organization formed on bases entirely idiosyncratic to a particular subject. The unit is defined solely by its repeated usage, and the experimenter may be totally unaware of the basis upon which it was formed.

The problem in specifying both types of organization is that the number of observed units is not readily interpretable. The number of observed units which are expected on the basis of chance, or which are possible at all, is a function of other characteristics of the recall performance. This means that when judging the degree of organization observed in a given subject's protocol, and especially when making comparisons between subjects or conditions, observed units of organization need to be interpreted in light of the other characteristics of the protocols from which they have arisen. Thus, observed units of clustering are to some extent a function of the total number of words recalled, the number of categories represented in recall, and the distribution of words recalled across the different categories. In the case of subjective, or inter-trial, organization the observed units are to some extent a function of the number of words recalled on each of the trials being compared, and the number of words common to recall on the two trials.

The general approach to dealing with this problem has been to find some type of derived score expressing the observed organization relative to the amount expected on the basis of chance, the maximum possible amount, or some combination of the two. There has recently been a proliferation of such scores (e.g., see Dalrymple-Alford, 1970; Dunn, 1969; Frankel & Cole, 1971; Hudson & Dunn, 1969; Pelligrino, 1971; and Roenker, Thompson, & Brown, 1971). A difficulty remains, however, in that each of these types of scores has somewhat different properties, and it is not yet clear which one controls best for the kinds of recall parameters which are involved here. Consequently, there has not been any determination of the single best measure, and different studies continue to utilize different measures. This, of course, raises concerns about the comparability of results and about the possibility of an integrated body of information. Continued investigation of the various measures and the possible consequences of specifying organization in these alternative ways seems, therefore, imperative.

Finally, the two forms of organization, category clustering and subjective organization, have almost without exception been investigated quite separately. This is obviously partly because the investigation of clustering began about ten years before that of subjective organization, but their concurrent investigation has been going on for about ten subsequent years. It is also true that they are operationally distinctive, but probably the most crucial difficulty has been that of establishing some common scale of measurement for the two forms. Since these are postulated to be two basic forms of organization in human memory, it would seem important to explore the relationship between them to determine, for example, whether one is used to the exclusion of the other or whether they interact in some specifiable way. Furthermore, there are the important questions of whether they display similar patterns of development with practice, are similarly affected by the same variables, etc.

In sum, the studies reported here all had bearing upon one or more of the following major questions:

1) The first question concerns the role of organization in free-recall verbal-learning and memory. There is a need for the continued investigation of the contention that organization facilitates retention, and the theoretical mechanisms by which it might have such effects.

2) Secondly, there is the question of the relationship between the two basic form of organization -- whether they are mutually exclusive or interact; whether they are similarly affected by the same experimental conditions, etc.

3) Thirdly, there is the continuing question about the conditions which are important in determining the extent to which organization will be utilized by the learner, i.e., what conditions help to maximize its use, etc.

4) Finally, there is the question of what type of score ought to be used to describe organization, and whether the type of score has important consequences for the conclusions reached in a particular experiment.

Experiment 1: The Role of Clustering in Free Recall

Purpose

It is frequently observed that more words are recalled from a categorized (C) list, or high conceptual similarity list, than from a non-categorized (NC), or low conceptual similarity, list (e.g., Cofer, 1967; Wood, 1968; Wood & Underwood, 1967: and Underwood & Freund, 1969). Since categorical clustering also occurs in recall from the C list it is possible to attribute the augmented recall of these materials to the use of this form of organization. These data would therefore appear to support the view that the amount that can be remembered is critically dependent upon the organization of recall (e.g., Mandler, 1967; Tulving, 1962a, 1968). However, Cofer (1967) has argued that the superiority of recall from C lists may not be dependent upon organization based upon category names. Wood and Underwood also suggest that the influence of high conceptual similarity occurs prior to the time of recall. Thus there is still a question about the role of clustering of the recalled responses in accounting for superior C list recall. The purpose of this study was to provide further evidence about the dependence of recall upon organization.

Method

A list of 18 numbers, chosen randomly from those between 1 and 50, was used for practice. Two separate C lists were prepared for the purposes of internal control and replication. Both C lists comprised 10 words in each of three taxonomic categories taken from the Cohen, Bousfield, and Whitmarsh (1957) norms. Two separate NC lists were also used. These lists were prepared by selecting 30 words so as to minimize obvious categorical relationships and duplications of first letters. The C and NC lists were also matched in terms of the frequency of occurrence (Thorndike & Lorge, 1944) and the mean length of the words. Five randomizations of each list were prepared.

A booklet technique was used for testing. Each S receivd a test booklet which presented, on successive pages, the instructions for free recall, the practice list items (numbers) written in a single column according to one of the randomizations, and a lined sheet for the recall of the practice list, one of the four experimental lists written in a single column, and a lined sheet for the recall of the experimental list. A period of 30 seconds was allowed for the study of the practice list and Ss were given 60 seconds for the recall of these items. The Ss were allowed to study their experimental list for 60 seconds, followed by a recall period of 120 seconds. The booklets were distributed in an order which systematically alternated between the four experimental lists. However, because of the nature of the planned analysis of the results approximately twice as many C list booklets as NC list booklets were distributed.

The Ss were 116 undergraduate students at Millersville State College. The data were collected in three separate classroom groups with the result that 40 NC list booklets were completed while 76 C list booklets were completed. The Ss turned out to be predominantly male, with only a few females in each condition.

Results

The results described here were pooled over the two separate experimental lists of each type because they were found to be quite equivalent. Additionally, the data on the amount recalled do not include extra-list intrusions or duplications of recalled items as words correctly recalled since analyses including these types of errors show the same outcome.

The mean number of practice list items recalled was broken down into those for the groups who later received the C and NC experimental lists. These data are shown in Table 1. The groups did not differ in the amount of material recalled from the practice list, F (1,114) = 3.71, p>.05. The mean number of words recalled from the two types of experimental lists is also shown in Table 1. The analysis showed that, overall, significantly more words were recalled from the C lists than from the NC lists, F (1, 114) = 30.31, p<.001, thereby replicating the results of previous studies.

The next step was to break the C list data down into that for Ss who showed a significant degree of category clustering in recall and that for Ss whose clustering did not exceed the amount which could be expected on the basis of chance. Each S was classified as a "clusterer" or a "non-clusterer" by determining whether, given the number of words which he recalled from each of the categories, the total number of units of stimulus category repetition (SCR) observed in his protocol was significantly different from that which would be expected on the basis of chance. The procedure for this was described by Puff (1963). Those 45 Ss whose observed clustering levels had chance probability values of less than .05 were designated as "clusterers" while those whose observed clustering did not exceed .05 under this null hypothesis were considered to be "non-clusterers". Table 1 shows the breakdown of C list data for these two sub-groups from that list.

The comparison of the number of practice list items recalled by the NC list Ss, the C list Ss who were significant clusterers, and the C list Ss who were not significant clusterers once again revealed no significant differences, F (2,113) = 1.99, p>.10. However, the analysis of the number of experimental list items recalled by each of these three groups did indicate the presence of significant variation, F (2,113) = 15.89, p<.001. Post hoc comparisons by the Scheffé method revealed that both of the C list groups ("clusterers" and "non-clusterers") recalled significantly (p<.001) more words

Table 1.

Mean Number of Items Recalled from the Practice List, the Non-Categorized List, and the Categorized List in Experiment 1

Subject Group	N	Mean Practice Items Recalled	Mean Experimental Items Recalled
All NC List Subjects	40	8.03	13.20
All C List Subjects	76	8.80	16.66
C List Subjects who were "Non-Clusterers"	35	8.94	16.26
C List Subjects who were "Clusterers"	41	8.68	17.00

than did the NC list Ss, and that there were no significant differences ($p<.05$) in the number of words recalled from the C list by the "clusterers" and "non-clusterers".

Discussion

These results seem to indicate that the high degree of recall from C lists, as opposed to NC lists, is not easily attributable to the occurrence of category clustering with this type of material. Even those Ss whose clustering did not exceed a chance level recalled significantly more words from the C list than were recalled from the NC list. Still further evidence for the lack of critical role in recall for organization in the form of clustering is provided by the observation that the "clusterers" did not recall any more from the C list than did the "non-clusterers". These results thus offer no support for the theoretical view that the amount that can be remembered is dependent upon the degree of clustering imposed at the time of recall.

It therefore appears that the general superiority of C list recall must involve some explanation other than clustering on the basis of broad conceptual categories at the time of output. This conclusion agrees, for instance, with Cofer's (1967) view that C list recall is augmented because of a higher level of interitem associative strength which increases the memorability of these stimuli, rather than because of organization based upon category labels. In the view proposed by Wood and Underwood, the greater recall from C lists is attributable to the occurrence of common implicit associative responses which result, via backward associations, in an increased frequency of elicitation of the representational responses for the individual stimulus words. Furthermore, this facilitation, or priming, is assumed to occur at the time the material is being studied, and prior to the time that it is recalled. Thus there are some viable alternatives to the explanation of the augmented recall with C lists in terms of categorical organization at the time of output.

Finally, it should be noted that a recent study by Thompson, Hamlin, and Roenker (1972) has shown evidence contrary to the results of the present experiment. In their study "high clusterers" recalled more words than "low clusterers". Some of the kinds of methodological differences which might be involved in the discrepant results include the fact that they did not employ a practice list to demonstrate that "high" and "low clusterers" were equivalent in basic ability; they used an index of clustering which did not provide exact probabilities; their lists comprised more categories and items per category; and their subjects were "primed" for the use of the categories. These discrepant findings indicate the need for further research to isolate which of the procedural variations might be responsible. For the time being Thompson, Hamlin, and Roenker's findings suggest a degree of caution in generalizing too broadly from the results of Experiment 1.

Experiment 2: An Investigation of Two Forms of Organization in Free Recall

Purpose

The organizational phenomenon of clustering, or the grouping of categorically related words in free recall, has been widely investigated since the original work by Bousfield in 1953. Similarly, the study of subjective organization, or the tendency of subjects to develop a relatively fixed order of recall across successive trials, has generated a great deal of interest and research since Tulving's work in 1962. A review by Shuell (1970) provides some indication of the scope and extent of investigation of these forms of organization.

Both of these forms of organization have to do with the ordering of responses in free recall, and both are postulated to be determinants of the amount of material that can be remembered. Thus it is surprising that there has been almost no direct comparative research and theorizing concerning the two types of organization. With the exception of some exploratory work by Mandler (1969), Roberts (1968), and Quaintance and Shapiro (1970) the study of clustering and subjective, or intertrial, organization has remained almost entirely separate.

Perhaps one of the reasons why the lines of investigation have proceeded separately lies in the definition of subjective organization. In his original paper on this phenomenon Tulving talked strictly in terms of subjective organization rather than using the designation of intertrial organization. He chose this terminology because of the "unrelated" nature of his stimulus words. While it is of considerable interest that such organization will develop even with unrelated words, this particular kind of stimulus material is not a necessary part of the operational definition. This form of organization refers, most simply, to agreement in the order in which the items are recalled on successive trials (i.e., to intertrial repetition). If this more generic view is adopted it is possible to measure this type of organization even for related, or categorized, words. Both clustering and subjective organization can then be determined in recall of the same materials. This leaves the problem, however, of finding a type of score, or scale of measurement, which will allow a meaningful direct comparison between the two types of organization. If that can be accomplished then it is possible to deal with questions about how the forms of organization may interact with each other, which type is more readily adopted by subjects, etc. The purpose of the present experiment was to explore some of these questions.

Method

The stimulus materials were two lists of 18 words. One of the lists was categorized (C list). This list comprised six words drawn from the taxonomic categories of animals, vegetables, and occupations in the Cohen, Bousfield, and Whitmarsh (1957) norms. The other list was non-categorized (NC list), or unrelated, in the sense that no obvious categorical relationships were intentionally built into it. The C and NC lists were matched on Thorndike-Lorge (1944) frequency of occurrence and mean number of syllables per word. Five randomized orders of presentation of both lists were used. The subjects were given standard instructions for free recall, and the words were presented by means of an automatic slide-projector at the rate of one every 2.5 seconds, with a period of 60 seconds being allowed for recall. Data for 15 presentation/recall trials were obtained from all subjects. Fifteen male undergraduate students at Franklin and Marshall College were presented with each list. They were assigned to lists in an alternating fashion on order of appearance at the laboratory.

Results

As expected on the basis of other research, the analysis of the number of words recalled indicated that significantly more words were remembered correctly from the C list than from the NC list, F (1,28) = 9.31, $p<.01$. Furthermore, the non-significant type of list x trials interaction indicated that the superiority was maintained uniformly across trials, $F<1$.

Both clustering and intertrial (subjective) organization scores were found for the recall of the C list. Only intertrial organization could be scored for the NC list. Observed amounts of both types of organization were then expressed as a percentage of the maximum possible amounts. This was done separately for each subject and the maximum possible values are a function of the specific parameters of his recall sequence or sequences. The mean percentage scores are shown in Table 2. It can be seen that the degree of clustering in the recall of the C list is strikingly greater than the degree of intertrial organization imposed in the recall of this same material. It can be seen that clustering began at about 40% and then rapidly approached the maximum possible amount while the observed amount of ITR never got higher than about 28% of the maximum possible amount. An analysis of the arcsin transformed values of these data, and excluding the clustering results for the first trials since ITR scores cannot be determined until the second trial, showed that the degree of clustering was significantly higher than that of intertrial organization, F (1,14) = 25.38, $p<.001$. The interaction of type of organization x trials was not significant, F (13, 182) = 1.34, $p>.05$ perhaps because the first trial clustering data had to be excluded. Another striking feature of the data shown in Table 2

Table 2.

Group Means of Observed Clustering and Intertrial Organization Expressed as Percentages of Their Maximum Possible Amounts in Experiment 2

	Percent Clustering	Percent Intertrial Organization	
Trials	C List	C List	NC List
1	39.84	Impossible	Impossible
2	57.84	10.67	12.89
3	76.92	11.04	12.65
4	77.63	16.88	8.34
5	80.76	10.50	12.08
6	82.98	18.32	13.04
7	92.79	16.05	11.76
8	92.07	14.51	13.01
9	93.34	25.52	19.62
10	94.30	17.28	14.38
11	96.39	18.92	18.63
12	86.19	16.74	25.56
13	99.48	18.71	27.64
14	96.34	25.88	24.44
15	92.00	25.86	25.70

is that the degree of intertrial organization imposed in the recall of the C and NC lists is so similar. The comparison of these data after arcsin transformation revealed no significant variation due to type of stimulus list, $\underline{F}<1$, and the type of list x trials interaction was not significant, $\underline{F}$ (13,364) = 1.62, p>.05.

The more common procedure of expressing observed organization as a deviation from the amount expected on the basis of chance was also employed in this study. The formulas for the expected amounts of both types of organization were taken from Bousfield and Bousfield (1966). Mean values of these deviation scores are shown in Table 3. It can be seen that the pattern of results is just about the same as that observed with the ratio scores, with the exception that in this case the first trial clustering results are reduced to the same level as those for intertrial organization. In the course of this work it was discovered that the maximum possible O-E deviation is much greater for ITR than is true for SCR. In other words, the two O-E scales have different end-points so that O-E (ITR) scores can get larger than O-E (SCR) scores can. In light of this, the use of O-E deviation scores would not seem to be a very appropriate way to compare the two types of organization. Thus, rather than placing any emphasis upon the slight discrepancy in results when using the ratio and O-E scores, the O-E score data were viewed as being less interpretable, but as generally supporting the conclusions to be derived from the ratio score comparisons.

A final analysis was performed to compare the C and NC lists in terms of the variability in the composition of the ITR units across trials, in order to determine the extent to which the subjects used the same or different ITR units across trials. A score was found for each subject by dividing the number of different observed ITR units (i.e., repeated pairs of words) which appeared in all of his protocols by the total number of his observed ITR units. The resulting mean values were .13 for the C list and .14 for the NC list. The application of White's Rank-Test yielded $\underline{T} = 213.5$ ($n_1 = n_2 = 15$), $\underline{p}>.05$. Thus, neither type of stimulus material led to a greater tendency for the consistent or repeated use of the same specific ITR unit pairs.

A Second Study

Since the findings of this study are somewhat limited because of the single set of stimulus materials, a second study was undertaken. The second study involved only C lists, but the strength, or potency, of the categories was varied. The lists comprised ten words in each of three categories, and each list was presented to a group of 14 subjects for a total of 15 trials. The mean frequency of elicitation of the category name by the category members according to the Cohen et al. norms was 28.7 for the "high" strength category list and 4.9 for the

Table 3.

Group Means of Observed Clustering and Intertrial Organization
Expressed as Deviations from Chance Expected Values
in Experiment 2

	O-E Clustering	O-E Intertrial Organization	
Trials	C List	C List	NC List
1	.34	Impossible	Impossible
2	2.02	.31	.42
3	4.69	.60	.47
4	5.23	1.35	.32
5	5.91	.65	.73
6	6.71	1.73	1.08
7	8.05	1.47	.70
8	7.97	1.35	1.01
9	8.15	2.84	1.95
10	8.69	1.89	1.30
11	9.00	2.07	1.90
12	7.04	1.79	3.04
13	9.55	2.20	3.42
14	9.32	3.36	2.70
15	8.61	2.82	3.26

"low" strength list. The results are shown in Table 4. It can be seen that, once again, the degree of clustering is substantially greater than that of intertrial organization, with virtually no influence being exerted by the strength of the categories.

Discussion

The results thus demonstrated a striking difference in the degree to which the two forms of organization were found in recall. The subjects employed organization in the form of relatively gross conceptual clusters to a much greater extent than they developed a fixed sequential ordering across trials. This is interpreted as reflecting the fact that the categorical relationships upon which clustering is based are seemingly more obvious than the assorted kinds of associations or relationships upon which units of subjective organization are based. That is, clustering can be assumed to be the result of the detection and utilization of a single common mediating response, or at most, a few common mediators (e.g., Bousfield, Steward, & Cowan, 1964). In contrast to this, the best evidence about the bases upon which ITR units are formed suggests that these may vary considerably even for the same subject (Abramczyk & Bousfield, 1967; Bousfield & Abramczyk, 1966). In other words, one unit might be formed on the basis of some idiosyncratic past experience, another on the basis of input contiguity, another on the basis of word length, etc. Thus, in order for subjects to impose comparable degrees of the two forms of organization it would be expected that they would have to engage in a greater amount and variety of mediating activities in the case of intertrial organization.

Another of the important results of this study is that the degree of subjective organization found in recall from the C and NC lists was virtually identical. The two kinds of lists were also quite equivalent in the variability in the composition of ITR units from trial to trial. These results strongly suggest that the two forms of organization are at least relatively independent. The occurrence of a great deal of clustering did not lead to any reduction in the amount of subjective organization as compared to that found with the NC list where there was not clustering. These results thus do not agree with Tulving's (1962a)speculation that items occurring in clusters have no fixed order and should therefore attenuate the amount of intertrial organization which depends upon just such a fixed order of recall.

The results of the second study suggest some generality for these conclusions, but there are other conditions under which different results might be found. For example, Mandler (1969) has found that some subjects do employ a great deal of serial ordering in recall of C lists when an incremental method of presentation is used. The incremental method involves presentation of only a single word on each trial, and under these circumstances some subjects seriate their recall in accord with the order of input. Mandler reports that this reduces

Table 4.

Group Means of Observed Clustering and Intertrial Organization
Expressed as Percentages of Their Maximum Possible Amounts
for High and Low Strength Category Lists
in the Second Study of Experiment 2

	Percent Clustering		Percent Intertrial Organization	
Trials	Hi-C List	Lo-C List	Hi-C List	Lo-C List
1	62.40	57.89	Impossible	Impossilbe
2	67.95	56.50	20.55	8.89
3	64.02	62.51	7.15	8.63
4	73.17	64.44	10.13	10.06
5	76.81	80.61	13.69	17.16
6	82.84	78.24	14.26	11.74
7	87.07	82.10	12.59	13.94
8	94.41	80.73	21.10	13.17
9	86.72	85.07	18.81	17.80
10	88.77	85.68	14.57	21.32
11	92.62	86.33	21.30	19.78
12	85.00	84.14	18.25	15.75
13	85.75	91.51	19.74	23.17
14	89.02	84.02	15.45	16.50
15	90.46	88.95	21.25	16.95

the degree of clustering. It might also be expected that since ITR is a "pair-wise" measure, the presentation of categorically related pairs of words, which would serve to cluster in recall, would serve to actually augment the degree of ITR. Similarly, it might be expected that with very large, broad categories, ordering within categories would become crucial and the degree of ITR might again be raised.

Experiment 3: Free Recall With Serial and Simultaneous Presentation of Categorized Word-Lists

Purpose

Only a small part of the data on human learning and memory performance has been collected under conditions where the subject has been given all of the items for simultaneous study. There is some evidence, however, that performance is facilitated by the simultaneous study method in comparison with the traditional serial method where the items are presented one-at-a-time. Greater recall after simultaneous study has been observed in one or more conditions of studies by Bower, Clark, Lesgold, and Winzenz (1969), Foote and Pollio (1970), and Winograd, Conn, and Rand (1971), as well as by comparing the results of Puff and Bousfield (1967) with those by Bousfield, Puff, and Cowan (1964). On the other hand, while there are apparently no cases where the serial method was found to be more advantageous, equivalent results for the two methods have been found in one or more conditions of experiments by Foote and Pollio, Heimer and Tatz (1970), and by Winograd *et al*. Thus, the simultaneous method seems more advantageous in some cases, but not always. The delineation of the factors involved in determining whether the simultaneous method will be superior or not has hardly begun.

The explanation of the superiority of the simultaneous method, when it is observed, is typically (see Bower *et al*., 1969; Foote & Pollio, 1970) that this type of presentation should lead to a greater degree of organization being imposed upon the material by the subject. That is, the simultaneous method affords a greater opportunity for the detection of any structure (associations, concepts, categories, etc.) in the material which the subject can use as a basis for organizing the material. Facilitation of recall with the simultaneous presentation should then follow from the greater degree of organization which can be developed with that method. One difficulty in evaluating this hypothesis is that the organization of recall has not been systematically investigated in the existing studies comparing the two kinds of methods. Furthermore, as indicated earlier, the simultaneous method has not always been found to be superior.

In the present research two further studies were conducted to determine the influence of serial and simultaneous study method upon the amount of material that can be remembered and upon the two different kinds of organization, clutering on the basis of conceptual categories, and subjective organization. These studies explored the importance of the kind of temporal equation between the two methods.

In most previous research the actual exposure time per item has been greater in the simultaneous condition than in the serial presentation condition. For example, if a stimulus list of 10 words is used, and if they are presented for a period of 20 seconds in the simultaneous condition, the average exposure time per item is 2 seconds. The standard procedure for the serial condition is to present the items at a 2 second rate. Because of the mechanical characteristics of equipment like slide projectors and memory drums some time is required to change from one stimulus to the next. Therefore, the average time that a stimulus is actually exposed for study is something less than 2 seconds with the serial presentation. Thus, it is possible that the superiority of the simultaneous method simply reflects an advantage in actual exposure time.

Exposure time has not previously been controlled, presumably on the assumption that average exposure time is not as important to control as is total processing time. Processing time refers to the amount of time available for postulated processes like perceptual registration, recognition, discrimination, rehearsal, organization, etc. Equating the two methods on the basis of average total processing time makes the assumption that subjects can engage in these processes just as effectively during the interstimulus interval as during the actual exposure of the stimuli. In light of the fact that if the two methods are equated on the basis of exposure time then the serial method involves a greater amount of processing time, most previous studies have controlled processing time.

Because of the possibility that quite different results might be obtained, the present studies compared the two methods under both types of temporal equation. Total exposure time was controlled in the first study while total processing time was controlled in the second study.

Method of the First Study

Fifteen stimulus lists comprising three taxonomic categories of 10 words each were constructed by drawing from a pool of six categories. In forming the lists, each category was used approximately the same number of times and appeared together in the same list with every other category approximately equally often. The words in the categories were drawn from the middle cultural frequency levels of the Cohen, Bousfield, and Whitmarsh (1957) norms. The categories were animals, clothes, furniture, occupations, vegetables, and weapons. Six randomized orders of presentation were prepared for each list. Thus, each input sequence contained a chance degree of grouping of words from the same category.

The same stimulus lists were presented by serial and simultaneous methods. In the serial condition, the words were presented successively by means of a slide-projector obtained from Lehigh Valley Electronics. This apparatus is essentially a Kodak Carousel projector modified so that an internal shutter can be controlled by automatic timing equipment. Each word was exposed for 2 seconds of study, and the interval between successive words was approximately one second (i.e., the total "change time" was 32.5 seconds instead of 30.0). In the simultaneous condition, all of the items were simultaneously available for study. They were presented by means of an American Optical Company overhead projector which was also controlled by automatic timers. The words appeared in a single column in the center of the screen, and they were exposed for a 60 second study period. The 60 second period in the simultaneous condition affords an average exposure of 2 seconds per word so that the time that the items were presented for study in the two conditions was equal.

All subjects were tested individually and were given standard instructions for multitrial free recall. Briefly, they were told about the mode of presentation, and that their task would be to write as many words as they could remember -- in any order in which they could remember them. They were given recall periods of 2 minutes, and a total of 12 trials was administered.

The subjects were 30 undergraduate summer-school students at Millersville State College who were paid for their services. They were assigned alternately to conditions on order of appearance at the laboratory.

Results of the First Study

The mean number of words recalled correctly in each condition is shown in Table 5 as a function of trials. These data were analyzed with conditions as a between-subjects variable and trials as within-subjects. The analysis indicated that while the serial method led to somewhat higher recall scores, the overall difference was not significant, F $(1,28)<1.0$, and the relationship between the scores for the two conditions did not change across trials, F $(11,308) = 1.44$, $p>.05$. Inspection of the values in Table 5, however, shows that recall with the serial method was consistently greater during the early trials. Individual tests on each of the first three trials were performed because of the rather obvious celing effect after a very few trials. However, none of these tests reached the .05 level of significance.

The tendency of the subjects to cluster related items together in recall was also examined. As usual, this type of organization was measured in units of stimulus category repetition (SCR), a unit of which was scored each time a word from any category was immediately followed in recall by another word from the same category. The amount of SCR expected by chance, E(SCR), was determined by the method described by Bousfield and Bousfield (1966). The two conditions were then compared in terms of the amount of observed clustering in excess

Table 5.

Mean Recall Measures Obtained with Total Exposure Time Equated between Serial and Simultaneous Methods of Presentation in the First Study of Experiment 3.

	Mean Words Recalled		Mean O-E (SCR)		Mean O-E (ITR)	
Trials	Serial	Simultaneous	Serial	Simultaneous	Serial	Simultaneous
1	16.00	14.73	5.53	2.68	*	*
2	21.80	10.33	7.12	5.06	1.36	.78
3	24.67	22.87	10.08	8.27	2.34	1.49
4	24.93	24.40	11.26	10.59	3.39	2.24
5	26.00	25.27	11.98	11.46	4.48	2.49
6	26.60	25.93	12.74	12.87	4.48	3.37
7	26.13	26.87	12.59	13.25	6.12	3.86
8	27.67	27.20	13.96	14.33	6.50	4.70
9	28.07	28.00	14.38	14.30	6.95	4.34
10	28.00	28.20	14.90	14.80	7.93	4.51
11	28.27	28.00	14.60	15.01	7.93	6.26
12	28.47	28.47	15.30	15.37	7.90	5.59

---

*Impossible

of the chance expectation, O-E(SCR). Mean O-E(SCR) scores are shown in Table 5. Inspection of these scores shows that those with the serial method are initially higher, but that there is a cross-over after 5 trials. However, the analysis of these scores showed the same pattern as for words-recalled--no significant effects except for the increase over trials. The result for study conditions was F $(1,28)<1$; that for trials was F $(11,308) = 73.29$, $p<.001$; and that for the interaction was F $(11,308)<1$.

Thus, somewhat surprisingly, the different types of presentation, or study conditions, had no effects upon amount of recall or upon either of the two types of measured organization. However, the various scores with the serial presentation were observed to be greater rather consistently. It is important to remember that while the actual duration of the exposure of the items was the same in the two conditions, the total duration of the input phase of a trial was longer in the serial condition because of the extra time between stimuli. It seems very possible that subjects make effective use of the time between stimuli for rehearsal. In other words, it is clear that the actual processing time was greater in the serial condition. Thus, it was concluded that whatever advantage the simultaneous method might have in terms of efficiency of processing, it was not sufficiently great to overcome the serial conditon's advantage in amount of processing time. Accordingly, in the next experiment equal amounts of processing time were arranged for the two conditions.

Method for the Second Study

The second study involved the same stimulus materials, apparatus, instructions, etc. The only change in the procedure was that in this case durations of the complete stimulus input phases of the two conditions were matahed. The temporal intervals in the serial condition were the same as those in the first study--2 seconds exposure per word and approximately one second between words, which sums to an acutal total duration of 92.5 seconds. Here, the input phase of the simultaneous condition was increased to the same total duration by giving a 92.5 second exposure of the list. Thirty additional subjects were recruited from the same source.

Results of the Second Study

The results obtained under these conditions showed precisely the same pattern as those in the first study. The data for each of the performance measures are shown in Table 6. As before, there were no significant differences between the study conditions in terms of number of words recalled, O-E(SCR) scores, or O-E(ITR) scores (F<1 in each case). Furthermore, practice had no differential effect on the two conditions, though the increase in all scores with practice trials was highly significant.

Table 6.

Mean Recall Measures Obtained with Total Processing Time Equated between Serial and Simultaneous Methods of Presentation in the First Study of Experiment 3.

	Mean Words Recalled		Mean O-E (SCR)		Mean O-E (ITR)	
Trials	Serial	Simultaneous	Serial	Simultaneous	Serial	Simultaneous
1	14.60	13.47	3.55	1.88	*	*
2	21.27	19.73	8.22	8.19	.85	.58
3	23.40	23.00	10.49	10.88	1.46	1.81
4	24.67	25.40	10.42	13.11	3.08	3.09
5	26.73	26.93	13.70	14.02	3.48	4.30
6	26.53	27.13	13.07	13.33	4.00	5.27
7	27.07	28.00	13.64	14.75	4.16	5.46
8	27.07	27.80	13.97	14.61	4.72	5.50
9	28.07	28.00	13.54	14.82	5.44	5.59
10	27.73	28.40	13.61	15.67	6.62	8.54
11	28.67	28.33	14.73	15.83	7.32	7.55
12	28.80	28.73	14.52	15.95	7.58	8.06

---

*Impossible

Discussion

The results of the two studies failed to show any superiority of performance with the simultaneous method regardless of whether exposure or processing time was controlled. This is especially surprising for the case where total processing time was controlled (i.e., in the second study) because this is the way in which previous experiments, some of which found positive results, were conducted.

These results illustrate once again that the simultaneous method is not invariably superior to the serial method. The important question still remains one of understanding what additional features of the study conditions determine whether or not the simultaneous method will augment performance. Several features of the present experiments may have been important in minimizing the differences between the methods.

In the first place, it appears that the present task was an easy one. Overall, the subjects in both conditions of both experiments were able to remember at least 75% of the material by the third practice trial. However, the improvement in recall with practice was highly significant so an absolute ceiling was not reached during the early trials, and it would have been quite possible to observe a differential effect under these circumstances if it was a robust one. Still, it seems that any potential advantages of the simultaneous method may be minimized where the task is otherwise an easy one.

Perhaps the nature of the stimulus materials also helped to minimize the magnitude of the effect. The materials presented here comprised three broad, and very obvious conceptual categories. One of the factors which Puff and Bousfield (1967) thought should lead to superiority of the simultaneous method was that it afforded the opportunity to detect various kinds of relationships among the words, presumably by allowing S to scan in varying directions, etc. It was assumed that the detection of such relationships, or associations, would facilitate the formation of higher-order organizational units. It seems quite possible that increased opportunities for detection of word-relatedness provided by the simultaneous methods was of little benefit in the present situation where the relatedness of the words was made very obvious. Some support for this interpretation is provided by the fact that the two methods led to the same amount of organization (which should reflect the formation of higher-order memory units).

According to this reasoning, then, the relative degree of influence of the method of presentation may depend upon the difficulty of detecting essential relationships in the material which will augment the formation of higher-order memory units. This suggests that Puff and Bousfield may have observed a large effect because they studied the recall of a word list involving a minimal number of normative relationships. The materials used by Bower, Clark, Lesgold and Winzenz (1969) employed lists that were hierarchically organized across four levels. It seems quite possible that, while this kind of list structure may be very beneficial in augmenting recall, it may also require finer discrimination at the time of stimulus presentation than does the detection of several very broad conceptual categories like those employed here. One study which may not be consistent with this tentative interpretation is that by Heimer and Tatz (1970). They studied the learning of lists of CVC nonsense syllables which should certainly embody a minimum number of associations and, according to the present reasoning, would therefore be expected to show a substantial effect of the method of presentation. However, such an effect was not observed. It is possible, of course, that nonsense materials behave differently from words, but is is also clear that the present interpretation of the data must be regarded only as tentative.

Another factor which may be involved in the discrepant results of the different studies may be the precise nature of the configuration in which the stimuli are presented. The two studies reported here and the one by Heimer and Tatz all employed a columnar configuration for simultaneous presentation. These are the instances in which no differences were observed. In contrast, the studies which found the simultaneous method to be superior utilized quite different arrangements. Puff and Bousfield used a circular array based upon two concentric circles with the words lettered in such a way as to minimize the ease with which they could be easily read in either a clockwise or counterclockwise direction. Somewhat similarly, Bower, Clark, Lesgold, and Winzenz used a hierarchical arrangement of the items. It seems quite possible that, as proposed by Heimer and Tatz, the subjects scan the simultaneous material serially when it is presented in a column. This would make the simultaneous presentation effectively identical to the serial presentation and would, of course, tend to minimize the performance discrepancy between the nominally different conditions.

Experiment 4: An Investigation of Memory Performance With Three Types of Presentation Sequences and Seven Measures of Organization

Purpose

Studies of input organization (IO) effects in free recall have most typically compared the influence upon amount recalled and conceptual clustering of what have been called blocked (BLK) and random (RND) arrangements of the same categorized word-list. In the BLK arrangement all of the items from one category, or concept, are presented before any instances of another category appear, and so on. Thus, the BLK list is completely organized in terms of the normative and experimenter-defined categories. The RND list, on the other hand, is generated by randomizing instances of the different categories, but most often the procedure also includes the restriction that no word can be immediately followed by another instance of the same category. The usual RND list, therefore, comprises a zero degree of organization in terms of immediate succession of words from the same category.

Comparisons of performance with these two types of input sequences have shown that the BLK arrangement virtually always leads to greater clustering of items from the same category in recall. However, the amount of material recalled has been found to be augmented by the BLK presentation in some cases (e.g., Cole, Frankel & Sharp, 1971; D'Agostino, 1969; Weingartner, 1964), but not in other studies (e.g., Elmes & Wright, 1970; Foote & Pollio, 1970; Yoshimura, Moely & Shapiro, 1971).

Other degrees, or levels, of IO have also been investigated. These have been specified according to the number of category repetitions (i.e., instances of immediate succession of words from the same category), average distance between words from the same category, or number of words from other categories intervening between words from the same category (Mayzner & Tresselt, 1961; Glanzer, 1969; Puff, 1966; and Shapiro & Bell, 1971). Amount recalled varied as a function of degree of IO in the studies by Glanzer and by Puff, and with high strength pairs (though not with low) in the study by Mayzner and Tresselt, and not in the study by Shapiro and Bell.

While the previous studies have covered a fairly broad range of IO levels, one type of manipulation of IO has not been investigated. A BLK arrangement of three items from each of three categories can be represented by the sequence AAABBBCCC, whereas a possible RND arrangement might be ABACBCBAC. It is a symstematically alternating (ALT) sequence like ABCABCABC which has not yet been studied. The question thus arises as to whether performance with the ALT list would be expected to be more similar to that with the RND or the BLK lists. Alternative predictions seem to follow from different theoretical views of the important factors involved in the explanation of IO effects.

Several theoretical views of IO effects have emphasized the basic importance of the degree of proximity of members of the same category (e.g., D'Agostino, 1969; Glanzer, 1969; Puff, 1966; Wallace, 1970, and Wood & Underwood, 1967). The consideration of the proximity of related items leads to the expectation of equivalent performance with the ALT and RND sequences, with both of these being inferior to the BLK sequence. In the example list of three items from each of three categories, the average distance between successive words from the same category is 0 intervening items with the BLK list, 2 items with the ALT list, and 1.5 items for the specific RND list given, but an average of 1.87 items over all possible RND lists of this type. The same expectation is reached when IO is specified in terms of the number of stimulus category repetitions (SCRs) in the input order. A SCR is scored each time a word from any category is immediately followed in the list by another word from the same category. Thus, the BLK list represents the maximum possible number of SCRs, or 6 in this case, while the RND and ALT sequences both involve 0 SCRs, or the minimum possible number.

On the other hand, the ALT list does involve a readily apparent sequential pattern. Thus, there seems to be a dimension of sequential structure along which the BLK and ALT lists are essentially similar, and both are superior to the RND list. This dimension will be referred to as sequential complexity, merely to distinguish it from the basic categorical structure of the list. This structure can be specified somewhat objectively, and at least partly independently of the proximity dimension, through the use of a descriptive coding (information) system similar to that described by Payne (1966a, 1966b).

Briefly, in Payne's work with binary and ternary digits, the coded form of a sequence was based upon the abstraction of repeated elements or series of elements. Applying a basically similar coding scheme to the same 9-word list example would ultimately yield codes of (A)3 (B)3 (C)3 for the BLK list, and (ABC)3 for the ALT list, but the best that could be accomplished for the particular example RND list would be (A)1 (B)1 (A)1 (CB)2 (A)1 (C)1, where the quantity inside the parentheses represents the largest possible sequential unit which is immediately repeated, and the number following the parentheses is the number of repetitions of the unit. Thus, much simpler codes are possible with the BLK and ALT lists, and this suggests, more generally, that these lists have a greater amount of potentially useful structure.

List structure information is postulated in several theoretical positions to play a major role in the explanation of IO effects in free recall. Knowledge about the structure of the list is thought to be important because it provides a basis for coding the material for storage and/or aids in the development of an effective plan for the retrieval of items from storage (e.g., Bower, Clark, Lesgold & Winzenz, 1969; Cohen, 1970; Lewis, 1971; and Newman, 1967). To the extent that these views imply that subjects may utilize any available type of structure, it would be expected on this basis that performance with the ALT and BLK lists would be similar, and both would be superior to the RND list.

Additionally, there is some evidence that this kind of sequential complexity is an important factor in several somewhat different types of tasks. For example, it has been found to be important in the recall of perceptual stimulus arrays by Glanzer and Clark (1962), and for subjects' judgments of the complexity of line-segment stimuli by Payne (1966a). Furthermore, a more general role for this kind of complexity in human memory has been suggested by Miller (1956) and Oldfield (1954).

Thus, the basic purpose of this experiment was to determine whether performance with the alternating sequence would be similar to that of the random order, as expected on the basis of the proximity measures, or whether it would be more similar to that of the blocked sequence, as suggested by the degree of sequential structure.

Method

The stimulus materials were those previously used by Puff (1966). A list of 15 numbers, randomly selected from those between 1 and 50, was used for practice. These numbers were presented to Ss as words rather than numerals. Five randomized orders of presentation of this list were used. The experimental materials comprised a list of 10 words from each of the taxonomic categories of animals, vegetables, and occupations. These items were chosen from the middle and lower frequency responses to the category names in the Cohen, Bousfield, and Whitmarsh (1957) norms.

The words were then arranged into BLK, RND, and ALT sequences. Fifteen separate lists of each type were prepared. In each of the BLK sequences all of the members of one category were presented successively before switching to the next category. The order of presenting the categories was counterbalanced across the multiple BLK lists, and the order of words within categories was randomized in every case. In the ALT sequences words from different categories were presented in a regularly alternating pattern, i.e., words from a given category appeared in every third serial position. The order of rotation through categories was counterbalanced across lists. The RND sequences were constructed by random sampling with the restriction that two words from the same category could never occur in immediate succession.

A summary of the average distance between words from the same category, the number of category repetitions, and the simplest possible sequential code for each type of sequence is shown in Table 7.

All stimulus items were typed in primary type in the center of 3 x 5 inch cards for manual presentation. A small flashing light, which was shielded from S's view, was used to pace the presentation.

The Ss were 45 male and female undergraduate students at Millersville State College who were paid for their participation. They were alternately assigned to one of the three input sequences in order of appearance at the laboratory.

All subjects were tested individually. The instructions emphasized that the order of recall was unimportant and that S could write the words in any order that seemed easy or natural for him. A single trial with the practice list was administered first. This list was presented at the rate of one item every 2.5 seconds, and a period of 2 minutes was allowed for recall. Then a single trial with one of the arrangements of the experimental list was conducted. These items were also presented at a 2.5 second rate, but 3 minutes were allowed for recall.

Results

The number of practice list items recalled correctly was analyzed first. As shown in Table 8, the mean recall from the practice list was somewhat higher for the ALT list Ss than for Ss in the other two groups. However, the analysis of these data indicated that there was no significant variation among the groups, $F(2,42) = 1.09$, $p > .05$. Even smaller differences were observed in terms of total items produced (i.e., including intrusions and duplications) in practice list recall, $F < 1$. Thus, it was concluded that there were no important differences among the three groups in basic ability for this type of task.

The mean number of experimental list words recalled correctly is also presented in Table 8. It can be seen that the mean correct recall from the RND and ALT sequences was virtually identical, and both were inferior to that with the BLK sequence. The overall analysis of these data indicated that this variation in recall scores was quite significant, $F(2,42) = 5.80$, $p < .01$. Once again, the same pattern was observed when the analysis was based upon data including intrusions and duplications, $F(2,42) = 5.44$, $p < .01$.

The degree of conceptual organization of recall, or the extent to which words from the same category were clustered together, after the different kinds of sequences was also investigated. Observed clustering was specified by the total number of stimulus category repetitions (SCRs) observed in a

Table 7.

Proximity Measures and Sequential Codes for the Three Types of Sequences Used in Experiment 4

	Proximity Measures		
Type of Sequence	Average Distance Between Members of Category	Number of Stimulus Category Repetitions	Simplest Possible Sequential Code
Blocked	0	27	(A)10 (B)10 (C)10
Alternating	2.00	0	(ABC)10
Random	1.94*	0	(CA)2 (BCA)2 (B)1 (A)1 (B)1 (CA)2 (B)1 (A)1 (CB)2 (A)1 (B)1 (A)1 (CB)2**

* This is the average over the 15 separate lists which were used; the range of values for the individual lists was 1.85 to 1.96.

** This is the code for one of the specific lists used; others have even less simple codes.

Table 8.

Group Means of Number of Items Correctly Recalled in Experiment 4

Sequence Group	Practice List Recall	Experimental List Recall
Blocked	7.40	13.87
Alternating	8.47	11.33
Random	7.60	11.33

S's recall protocol. A SCR was scored each time a word from any category was directly followed in recall by another word from the same category.

Because the values of observed clustering, O(SCR), which can be expected on the basis of chance, or are even possible at all, are a function of properties of the recall sequence such as total number of words recalled, number of categories represented in recall, and the distribution of recalled items across categories, the proper interpretation of O(SCR) data requires that these other recall parameters to be taken into account. This is accomplished by forming some type of derived score which expresses O(SCR) relative to chance expectations and/or maximum possible values.

Previous studies of input organization effects have almost exclusively used either a slightly modified form of Bousfield's (1953) ratio of repetition (RR) or Bousfield and Bousfield's (1966) O-E (SCR) deviation score. Both measures have recently been heavily criticized by several investigators (e.g., Dalrymple-Alford, 1970; Dunn, 1969; Frankel & Cole, 1971; Hudson & Dunn, 1969; Roenker, Thompson & Brown, 1971) on a number of statistical grounds. For example, the RR measure does not involve an accurate estimate of maximum possible O (SCR) values and does not take chance into account directly. The O-E(SCR) score expresses the deviation from chance expectations, but does not consider maximum possible values so that it varies substantially with the amount of recall. While these measures may be satisfactory for describing clearly how much clustering occurred in a single recall sequence, the inherent problems make it difficult to make comparisons across different conditions or trials where parameters as total recall are likely to vary.

Each of the critics has proposed a new type of clustering score which is postulated to be more readily interpretable than all of the others. For instance, Roenker et al. argue that the deviation of observed from chance clustering should be viewed relative to the maximum possible deviation from chance. Dalrymple-Alford suggests that the observed deviation from chance be considered relative to the entire range of possible O(SCR) values, i.e., from minimum possible to maximum possible. Dunn, Hudson and Dunn, and Frankel and Cole feel that the best procedure is to express clustering in terms of standardized scores by taking the deviation of observed from expected clustering relative to the standard deviation of such values. Unfortunately, there has not yet been any agreement upon the best single score, nor any compelling way to decide between them. Analytical arguments and artificial experiments have simply served to demonstrate that all of the measures have both merits and apparent weaknesses.

Because of the uncertainty about the best procedure to use, as well as the need for more comparative study of the various measures, the present data were analyzed by both of the measures used in most previous studies of input organization effects, as well as by five of the newer measures. Formulas providing somewhat fuller definitions of all measures are given in Appendix B, though the details of their calculation must be found in the original articles. Mean values of all measures for each of the three types of input sequences are shown in Table 9 along with the results of the analyses of variance.

The overall analyses of variance were followed in each case by the application of post hoc comparisons by the Scheffé method. The pattern of results was the same with every measure. No significant differences were found between the RND and ALT sequences, but both of these were inferior to the BLK list.

Discussion

The results indicate that both recall and clustering were quite equivalent with the RND and ALT types of input sequences and that both of these types were greatly inferior to the BLK sequence. These results thus conform to the predictions made on the basis of the proximity of items from the same category, regardless of whether proximity is specified by the average number of words intervening between two words from the same category or in terms of the number of SCRs present in the input list.

All of the theoretical views emphasizing simple proximity are thus supported and there are no apparent grounds for distinguishing between various different ideas of why proximity is important. Thus, it could be because of any of a number of more specific processes or mechanisms. Puff (1966) proposed that both amount of recall and clustering should be a function of the distance between related words because of an influence upon the availability, or priming, of common implicit associative responses (IARs). That is, since priming effects are assumed to dissipate with time, the more contiguous the instances of priming the same IARs, the greater the summation of priming effects, and the more available the common IARs as effective mediators among the category members. Wood and Underwood (1967) espoused a somewhat similar view, but put the emphasis upon the priming of representational responses and the subsequent availability of the individual words. Glanzer (1969) stressed that distance was important because it determined the extent to which related items would be short-term storage (STS) at the same time. He further suggested that the simultaneous presence of related words in STS simplified and facilitated the rehearsal process, and thereby increased the probability that the words would enter long-term storage (presumably together). Somewhat similarly, D'Agostino (1969) proposed that the distance between related words affects the effective amount of time for processing those words together as a single unit. That is, in order to process, integrate, or rehearse them together, S must recall the previously presented word when the later word is presented. As

Table 9.

Group Means of Various Clustering Measures and Results of Analysis of Variance for Each Measure in Experiment 4

Clustering Measure	Means for Sequences Blocked	 Random	 Alternating	F(2,42)	p value
Ratio of Repetition (Bousfield, 1953)	.64	.40	.46	10.49	<.001
O-E (SCR) Deviation Score (Bousfield & Bousfield, 1966)	4.56	1.11	1.95	13.50	<.001
Adjusted Ratio of Clustering (Roenker, Thompson & Brown, 1971)	.60	.18	.32	8.74	<.001
D_A Index Scores (Dalrymple-Alford, 1970)	.38	.12	.20	7.83	<.005
SCR Standard Scores (Hudson & Dunn, 1969)	2.72	.75	1.30	11.31	<.001
Standard-Score Runs-Test (Frankel & Cole, 1971)	-2.72	-.75	-1.30	11.31	<.001
D_N Index Scores (Dunn, 1969)	2.65	.85	1.38	12.11	<.001

the distance between related items increases, it is assumed that it should take more time to search for and recall the earlier item, thereby reducing the amount of time which can be devoted to effectively processing the related items together. Finally, Wallace (1970) has also asserted, without speculating as to the mechanisms involved, that items which are experienced contiguously will tend to be recalled contiguously.

Alternatively, these results are not consistent with the predictions made on the basis of the degree of overall list structure, and thus do not seem to support those views of input organization effects which stress the utilization of list structure as a basis for coding the material for storage or for the development of systematic plans for the retrieval of material from storage (e.g., Bower *et al.*, 1969; Cohen, 1970; Lewis, 1971; and Newman, 1967). The basic problem here, in short, is that the ALT sequence involves a great deal of objectively specifiable sequential structure which should have afforded some advantage over the RND list if *Ss* do indeed primarily base storage or retrieval strategies upon list structure.

There is a possibility, of course, that the sequential structure of the ALT list was not utilized effectively because it was not even detected by the *Ss*. There is no way to refute such a possibility in the present experiment, but there is a fair amount of evidence that the point in time at which the detection of basic categorical structure occurs is not crucial. No reduction in IO effects has been found even when *Ss* have been given quite detailed information about the categorical composition of the list prior to presentation (e.g., D'Agostino, 1969; Newman, 1967). Furthermore, Cohen (1970) found that manipulation of the distance into the input list before the categorization could be detected did not produce any differences in total recall. However, the demonstration that *Ss* are in fact sensitive to the presence of this kind of structure would further strengthen the present conclusions, and would be an interesting contribution in its own right.

The similarity of the results with all of the different clustering measures leads to several important conclusions. First of all, the findings of previous studies showing greater clustering with the BLK list than the RND list using the earlier measures are supported. Even the more recently proposed, more sophisticated, measures show the same pattern of results. This strongly suggests that the greater clustering with the BLK list cannot be regarded as a statistical artifact arising, for example, because there is greater recall from the BLK list, etc. Apparently, the degree of IO influences the tendency to group conceptually related words quite directly. This finding also lends confidence to the conclusions of earlier clustering studies in general, for the issue which is raised by the emergence of the new measures is not only which to use in future research, but also how much the results of earlier clustering studies can be trusted.

The present results do not offer much help in choosing a single measure for subsequent work. In fact, what is clear is that, under some circumstances at least, the choice of measures is not very crucial. However, it is also the case that the present data did not involve a very wide range of different recall parameters, and thus would not be really optimal for producing discrepant results with the different measures. Until research can identify the single best measure, the present strategy of using a number of different ones would seem to be recommended.

Experiment 5: Temporal Properties of Organization in Recall of Unrelated Words

Purpose

One of the major endeavors in the study of the organization of human memory has been to be able to specify the units or structure of the organization. One of the most widely used techniques is the specification of subjective organization (Tulving, 1962a).

Subjective organization refers to the observation that, even when "unrelated" words are used, subjects tend to recall sequences of items in the same order on repeated trials. Following Bousfield, Puff and Cowan (1964) and Bousfield and Bousfield (1966), the individual units of subjective organization are pairs of words which are repeated in the same sequential order on two successive trials. These units are called intertrial repetitions, or ITR units.

Tulving reasoned that such an event represents a reflection of a memory unit because, if the subject is free to recall the words in any order, and if he repeatedly recalls some words together in contiguous positions in the output sequence, then the words are behaving as if they have become a single unit. Thus, part of the appeal of the study of this type of organization is that the investigator does not have to be able to specify the bases upon which Ss form these units. On the other hand, this also means that the observed units may not indicate to the investigator any immediately obvious basis for their formation. For example, if the subject repeats a pair like table and hippopotamus on two successive trials, it may be difficult to believe that these words really constitute a single functional unit in the subject's memory. Furthermore, there is the concern that this kind of "pairwise" measurement cannot possibly reflect accurately the "real" nature of the organization which Ss use.

Thus, the primary purpose of this study was to provide some further evidence about the functional "reality" of measured units of subjective organization. The basic assumptions here were the same as those proposed by other investigators (e.g., Mandler, 1970; McLean & Gregg, 1967; Pollio, Richards & Lucas, 1969); namely, that free recall performance is a reflection of the nature of storage in memory, and that items which are emitted in close temporal sequences represent some kind of a unit in memory. Here, units defined on the basis of repeated contiguous recall were examined to see if they showed close temporal contiguity as well.

Method

The stimulus materials were two lists of 15 words which were unrelated in the sense that obvious taxonomic categories and duplications of first letters were avoided. The words covered a range of Thorndike-Lorge (1944) values from one to AA, and the two lists were matched on mean frequency, as well as mean number of letters per word.

The subjects were tested individually by means of a booklet technique. The first page of the booklet contained instructions to the effect that there would be a series of recall trials, and that the subject was free to recall the words in any order each time. There were also some cautions about things like extraneous noises since the recall was oral and was tape-recorded. Subsequent pages of the booklet alternately presented the words and instructed S to recall them. Word presentation pages showed the words in a randomized order in a single column which the subject was allowed to study for 30 seconds. Periods of 30 seconds were also allowed for recall. A tone was used for pacing, and a total of 15 trials was administered.

The subjects were approximately equal numbers of male and female undergraduate students at Millersville State College who were paid for their services. They were assigned alternately to the two lists. Five subjects had to be excluded from the analysis because of extraneous noises which occurred during the session or failure to follow instructions. Data are thus presented for a total of 39 subjects--20 for one list and 19 for the other.

Two basic kinds of data were extracted from the tape-recorded recall sessions: the record of the words recalled, and the interword times, or latencies, from the termination of each word to the onset of the next word. The times were obtained by processing the tapes through an apparatus comprising the following components: a Krohn-Hite bandpass filter to limit the signal to the upper voice-range; a Grason-Stadler voice-operated relay set with threshold values equal to 30% changes in signal amplitude; a digital-logic signal-transducer to provide a signal with constant amplitude and varying duration; and the pen on a Beckman Type R dynograph. The net result was that each word was represented as a square-wave form, and the interword latencies could be read in hundredths of a second. The estimated error in the system was about 3%.

Results

The results are pooled over the two replication lists. The inclusion of lists as a separate factor in analyses indicated minimal differences, with some important exceptions to be noted. All results are presented as a function of positions in the recall sequence. Because of variation in the number of words recalled across subjects and trials, every recall sequence was partitioned into five segments, or stages, by a "Vincent-like" method similar to that used by Bousfield and Cohen (1953). This procedure differs from the original Vincent method in two basic respects. In the first place, the procedure was used here to equate recall sequences of different lengths, rather than trials-to-criterion sequences. Secondly, scores were always placed entirely into one of the segments according to predetermined rules, rather than sometimes being split proportionately between two segments as in the original Vincent method. Additionally, the data for the fifteen trials were collapsed to mean scores for each of three trial-blocks. Because of the requirements of the measurement of organization, all analyses involved blocks comprising trials 2-6, 7-10, and 11-15. Thus, the analyses show what happens across five successive stages of the recall sequences within trials, and what changes occurred across three blocks of practice trials.

The first treatment involved all of the interword times completely ignoring any organization which was taking place. Group means of individual subjects' median times are shown in Table 10. The significant effects in the analysis of these data were: fifths of output, F (4,184) = 155.24, $p<.001$; trial-blocks, F (2,74) = 22.20, $p<.001$; and the fifths x trial-blocks interaction, F (8,296) = 27.52, $p<.001$. Post hoc comparisons by the Scheffé method indicated that across stages of output within trials there was an initial period of fairly consistently rapid recall, with the interword times increasing sharply only in the later part of the recall sequence. The number of individual words represented in the initial flat portion of the curves was approximately 4-5 in early trials, and 8-9 in later trials. Practice served to reduce only the times at the end (the final fifth) of the recall sequence and had no detectable effect upon the times for those words given earlier in recall.

Subsequent analyses took subjective organization into account. Following the Bousfield and Bousfield (1966) procedure, a unit of intertrial repetition, or ITR unit, was scored each time two words recalled consecutively on one trial had also been recalled consecutively on the previous trial.

First, the relative amounts of organization at each of the five stages of the recall sequence within trials were examined. Density of organization was expressed by taking the number of observed ITR units as a percentage of the number of opportunities. Mean density of organization functions are shown in Table 11. The only significant effects revealed by the analysis of these data were: fifths of output, F (4,148) = 3.77, $p<.01$; the fifths x lists interaction, F (4,148) = 3.92, $p<.005$; and trial-blocks, F (2,74) = 29.15, $p<.001$. The fifths x lists interaction indicated simply that, while the fifths effect was significant in both lists, it was more highly significant in one list. Post hoc comparisons indicated that only the first

Table 10.

Group Means of Subjects' Median Interword Times (Without Regard to Organization) for Five Stages of the Recall Sequence Within Trials, and for Three Blocks of Trials, in Experiment 5

Trial Blocks	Successive Fifths of Output 1	2	3	4	5
1	.27	.50	.60	1.14	3.19
2	.33	.64	.76	1.11	2.05
3	.34	.64	.71	.93	1.40

Table 11.

Mean Density of Observed ITR Units as a Percentage of Opportunities. Results are for Five Stages of the Recall Sequence and for Three Blocks of Trials in Experiment 5

Trial Blocks	Successive Fifths of Output 1	2	3	4	5
1	14.85	13.07	13.97	14.27	10.38
2	27.12	22.64	22.53	24.66	17.19
3	40.05	31.10	31.01	32.32	28.05

and final fifths of the output sequence were significantly different from each other. Thus, the density of organization decreased quite gradually across successive stages of output. Practice markedly increased the overall level of organization, but did not modify the distribution of units across stages of the recall sequence (i.e., the fifths x trial-blocks interaction was not significant).

For the next analyses, all interresponse times were classified as being between two words comprising an intertrial repetition unit (called ITR times) or between two words which did not comprise an ITR unit (called non-ITR times). Some problems of data analysis were encountered at this point. Only four subjects had both ITR and non-ITR times in every fifth of the recall sequence in all three trial-blocks. Since the variables of interest were within-subjects, an overall analysis including all of the factors simultaneously was precluded. However, several kinds of separate analyses were performed.

Initially, a test was conducted on scores found by averaging over both trial-blocks and fifths of output. The mean ITR time was .62 second while that for non-ITR times was 1.32 seconds. The test between the paired scores indicated that the overall difference was quite substantial, $\underline{t}$ (38) = 12.13, $\underline{p} < .001$. This afforded some protection for the following analyses.

In one case, results were tabulated for those subjects who had at least one ITR and one non-ITR time in every fifth of a given trial-block. This provided data for three separate analyses--each one representing a comparison of ITR and non-ITR times across fifths of output within one of the trial-blocks. These data, along with the number of subjects represented in each case are shown in Table 12. Each analysis included fifths and type of unit (ITR vs. non-ITR) as within-subject variables. As shown in Table 13, the results of each trial-block were very similar, indicating significant effects of units, and of fifths, as well as significant units x fifths interactions. These results, combined with more specific comparisons, thus revealed that ITR times were significantly shorter than non-ITR times, but only after the initial two fifths of the recall sequence in the first two trial-blocks, and after the initial three fifths in the third trial block. Again in terms of individual words, the ITR times were significantly shorter only after the subject had recalled approximately five words in early trials and six words in later trials. It can also be seen that the discrepancy between ITR and non-ITR times arises primarily because non-ITR times increased substantially after the first few fifths of output, while ITR times were considerably less affected, though the relatively slight increase in these times at the very end of the sequence was also significant. Since it was not possible to compare across these sets of data to test for practice effects, a second set of analyses was performed.

Table 12.

Means of Subjects' Median Interresponse Times for ITR and Non-ITR Units at Five Stages of the Recall Sequence and for Three Trial Blocks in Experiment 5

Trial Blocks	N	Type of Unit	Successive Fifths of Output				
			1	2	3	4	5
1	14	ITR	.30	.55	.41	.39	1.00
		Non-ITR	.34	.46	.71	1.72	3.37
2	11	ITR	.27	.52	.48	.39	1.24
		Non-ITR	.36	.67	.89	.92	2.80
3	23	ITR	.39	.48	.64	.51	1.28
		Non-ITR	.46	.76	.95	1.76	2.36

Table 13

Results of Analyses of Interresponse Times
Across Fifths of Output Within Each Trial Block
in Experiment 5

	First Block		Second Block		Third Block	
Source	*F*	*df*	*F*	*df*	*F*	*df*
Units	54.24	1,13	20.37	1,10	22.86	1,21
Fifths	31.94	4,52	17.50	4,40	10.96	4,84
U x F	22.53	4,52	6.24	4,40	5.87	4,84

Note.--All *p* values <.005 or .001.

The data for the analysis of practice effects were obtained by finding those subjects who had both ITR and non-ITR times in all three trial-blocks for a given fifth of the output sequence. In this case therefore, there were five separate analyses--each one testing for trial-blocks effects within one of the fifths of output. The data in this form are shown in Table 14. This set of analyses revealed that there were no significant effects of practice, except in the middle fifth of output, $F\ (2,34) = 5.42$, $p<.01$, but this effect was unexplainably attributable to only one of the two stimulus lists. None of the trial-blocks x units interactions were significant, so it is clear that ITR and non-ITR times were not differentially affected by practice at any stage of the output sequence.

A further analysis of the effects of practice focused upon changes in interresponse times when the same ITR units were repeated varying numbers of times. Tabulations were made of all of the instances where the same unit was repeated at least two times, at least three times, and so on through ten times. These data along with the number of instances of each frequency of repetition are shown in Table 15. Separate analyses for repeated measures were performed for each frequency of repetition. Only the units which were repeated at least twice showed a significant decrease in time with repeated use, $F\ (1,37) = 5.35$, $p<.05$. In spite of the impression given by some of the absolute values in Table 15, times did not decrease significantly (all $ps>.05$) with any higher frequency of repetition. Because of the high degree of variability, these results were checked by the application of Friedman's nonparametric analysis by ranks, and in this case not even units repeated at least two times were significant.

Discussion

The major implication of the present results is their support for the "reality" or "validity" of measured units of subjective organization. It was found that words which are defined as being organizationally linked by the examination of recall ordering also show shorter interword latencies than words which the repeated ordering measure says are not organizationally related. However, the temporal differences between ITR and non-ITR units did not appear until after the subjects had recalled a number of words which is approximately equal to the immediate memory span. More specifically, those words recalled early in the output sequence were emitted rapidly, regardless of organizational links or their absence. It was during the later stages of the recall sequence, when the immediately accessible items were largely exhausted, that the significant effects of organization took place.

The functional "reality" of these ITR units is further supported by the similarity between their temporal characteristics and those found with units defined in other ways. Mandler (1970) reported basically similar results for units defined by the categories into which subjects sorted the words before

Table 14.

Means of Subjects' Median Interresponse Times for ITR and Non-ITR Units for Five Stages of the Recall Sequence and Three Blocks of Trials in Experiment 5

	Successive Fifths of Output														
	1			2			3			4			5		
	N=18			N=19			N=19			N=24			N=19		
Trial Blocks:	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
ITR Units:	.30	.27	.36	.74	.51	.36	.43	.66	.58	.64	.67	.72	.81	1.03	.88
Non-ITR Units:	.32	.45	.47	.58	.65	.66	.69	.92	1.11	1.59	1.20	1.37	2.82	2.67	2.88

Table 15.

Mean Interresponse Times in Seconds between Words in Units Which Were Repeated at Least X Times in Experiment 5

Instance of Use	Times Unit was Repeated								
	2	3	4	5	6	7	8	9	10
1	.66	.57	.67	.73	.83	.74	.36	.32	.31
2	.51	.44	.42	.45	.54	.50	.61	.58	.71
3		.39	.39	.39	.45	.45	.53	.53	.57
4			.46	.52	.51	.46	.54	.57	.53
5				.43	.47	.40	.45	.36	.29
6					.40	.40	.47	.41	.43
7						.41	.38	.30	.29
8							.34	.29	.25
9								.34	.49
10									.29
N	38	34	31	25	18	16	14	11	9

recall. There is also striking agreement with the interresponse times found by Pollio, Richards and Lucas (1969) with units defined on the basis of clusters of words from the same experimenter-defined taxonomic categories. Their results show virtually the same kind of interaction between the effects of organizational units and successive stages of the recall sequence.

A second major aspect of the results concerns the effects of practice. Neither ITR nor non-ITR times, at any stage of recall within a trial, were much altered with increasing practice. Furthermore, after a decrease in latency the second time the same ITR unit was repeated, additional repetitions of a unit did not result in any further reduction in time between the members of the unit. This lack of practice effects seems entirely reasonable for non-ITR times, but the weakness of the effect upon ITR times is a bit surprising. It was thought that they would continue to become stronger, or better consolidated, with repeated usage. However, because of some possibility of unknown item-selection effects in these analyses, definite conclusions here will require additional direct experimental manipulation.

Practice did, however, have a significant effect upon interword times considered without regard to organization. The longer times at the end of the sequences were seen to decrease substantially. Since changes in ITR and non-ITR times did not occur, it appears that the explanation of this effect involves two other factors. It is at this same place in the output sequence that ITR times are appreciably shorter than non-ITR times, and with increasing practice the number, or density, of ITR units increases. That is, with increasing practice a larger proportion of the later interword times are short ones from ITR units, and the average time at the end of the sequence comes down. Practice also produces increasing numbers of ITR units at the beginning of the recall sequence, but all of the first words recalled are emitted so rapidly that the increasing density of ITR units has no detectable influence upon average interword times for those words given early.

Incidentally, it should be noted that the present results for interresponse times without regard to organization are basically similar to those recently presented by Murdock and Okada (1970) for a single-trial recall over a series of lists. One apparent difference in the two sets of results is that the present functions remain essentially flat further into the recall sequence than is seen in the Murdock and Okada data. This probably does not reflect the different ways in which the data were partitioned. This was checked by plotting the present first trial data in the manner used by Murdock and Okada with no appreciable change in the pattern. Perhaps there is another kind of practice effect which arises from experience with multiple lists.

Another of the important findings here was the distribution of ITR units across successive stages of the recall sequence. The density of organization began at peak level in the initial part of the recall sequence, and then decreased quite gradually until the lowest density occurred at the end of the sequence. Also, while practice raised the overall level of organization, it did not change the relative distribution of units.

At least a rough comparison can be made between these data and those for organization in the form of clustering by Bousfield and Cohen (1953), though any of a number of procedural features of either study could have influenced the nature of the density functions in addition to the differences due to the type of organization studied. Their stimulus materials comprised four taxonomic categories of fifteen words each, they used a serial rather than a whole method of presentation, their subjects wrote their recalled words, and they obtained data from only five trials. The density of clustering as a function of Vincentized tenths of the recall sequence was presented for each of the five trials. Their functions look very different from those obtained here for subjective organization. Specifically, the overall level of clustering is higher; practice does modify the shapes of the functions, operating primarily to raise the initial segments; and within a few trials, the functions assume a quite negatively accelerated form. About the only similarity that can be seen is that, after the first few trials with the categorized list, there is a general tendency for both types of organization to decrease monotonically across successive stages of the recall sequence. The reasons for the discrepancies are far from clear. Perhaps, for one thing, clustering at the beginning of the recall sequence is more disrupted, in the first few trials, by the subjects' tendency to begin their recall with words from the favored serial positions. With a randomized order of presentation, this would be likely to produce an initial run of words from different categories. Perhaps with some practice, subjects learn to abandon the use of a strong serial-position strategy in favor of the fuller completion of the category with which they begin recall. In the case of subjective organization, the tendency to begin recall from favored serial positions may not have such disruptive effects because the organization does not depend upon a few broad experimenter-defined categories, and in fact, the order of presentation in early trials may actually be the source of some of the units that the subjects form.

Finally, the restriction in this experiment to the study of units of organization defined only as ordered pairs of items (i.e., unidirectional ITR units) undoubtedly underestimates the amount of organization actually imposed by the subjects. Unordered pairs (bidirectional ITR units) or larger clusters of words, for example, might certainly constituee equally "real" units, and would thus be expected to be separated by similarly short interresponse times. The classification of these other kinds of potentially short latencies as non-unit times in the present study makes the observed effects of ITR units even more convincing.

Experiment 6: Effects of Instructions Upon Recall and Clustering with Conceptually Categorized Materials

Purpose

A number of recent free-recall studies of human memory have involved the instructional manipulation of the organization of the material by the Ss and the examination of the consequent effects upon the amount of material recalled. The interest in this type of study stems primarily from the reasoning that this represents a relatively straightforward way to test for the extent to which the amount of recall is critically dependent upon the organization imposed upon the material by the learner (e.g., Mandler, 1967; Tulving, 1962a).

The majority of the evidence in support of the dependence of recall upon organization is correlational. That is, it is virtually always observed that both organization and recall increase over successive practice trials (e.g., Bousfield, Puff & Cowan, 1964). It has also been sometimes found that Ss who organize more also show higher recall scores within trials (e.g., Tulving, 1962a). Regardless of how frequently these kinds of observations are made they still involve the same basic inability to infer cause and effect that is inherent in all correlational evidence.

Tulving (1962b) reasoned that one way in which a more direct type of evidence can be obtained is by experimentally inducing an increase in organization and observing any resultant effects upon the amount of recall. If the amount of recall is augmented by the induced increase in organization, then the dependence of recall upon organization is more clearly demonstrated. Tulving further proposed the use of instructions as a way of promoting different levels of organization while affording the control of having all Ss recall exactly the same materials. An essential part of this kind of manipulation is , of course, that it is in fact possible to significantly increase the amount of organization by instructing Ss to this effect.

Tulving's (1962b) original study involved a list of "unrelated" stimulus words and subjective organization, or the tendency of Ss to recall a list of words in the same order over a series of practice trials. After the first few trials, one group of Ss was allowed to continue under standard instructions deemphasizing the importance of order, whereas a second group was then instructed to recall the words in alphabetical order each time. The alphabetical ordering group subsequently showed better recall, and it was concluded that this was due to the increase in organization in the form of fixed sequential (alphabetical) ordering. However, Earhard (1967) later demonstrated that a fair part of the effect of alphabetical ordering is due to the fact that the letters of the alphabet serve as cues for the recall of the words in the list which start with those letters rather than to the fixed sequential organization per se.

Later studies by Mayhew (1967) and Puff (1970) avoided this confounding somewhat by not stipulating that alphabetical ordering had to be used. The Ss that were instructed to organize in these studies were given a number of different possible strategies (e.g., making up stories, grouping by images, etc.) which they could use to develop a high degree of fixed sequential organization. The two studies agreed in finding that Ss instructed to recall in a fixed order did indeed show a significantly greater amount of this type of organization, but these Ss recalled no more words than those given standard instructions.

A developmental study by Hultsch (1969) further confirmed that instructing Ss to organize in some way produced no greater recall than standard instructions, but superior recall with instructions for alphabetical ordering was once again observed. These conclusions held equally well for Ss who were respectively 16-19, 30-39, and 45-54 years old. Bosner (1971) compared performance by children in grades 1, 5, and 9. In her study, standard instructions were contrasted with others asking the Ss to rehearse each item as it was presented by repeating it twice out loud. The results indicated that there was a significant developmental effect in this case. It was only in grade 5 that any substantial facilitation of organization and recall resulted from the instructions to find mediational links, though it was also found that the instructions for forced repetition had a detrimental effect upon performance in the 9th grade group.

A slightly different kind of design, but one producing further evidence about the relationship between instructions, organization, and the amount of recall, is that used by Mandler (1967). Basically, Mandler found that Ss instructed to categorize a list of words were subsequently able to remember as many of the words as Ss explicitly instructed to try to remember them. Mandler concluded that since instructions to categorize are equivalent to instructions to remember, this supports the view that categorizing (organizing) is the crucial thing that Ss do when asked to remember verbal materials. Some generality for these conclusions has been provided by Sturges, Crawford, and Nelson (1971) as well as by Nelson, McRae, and Sturges (1971) though their work has also suggested some limitations upon these findings when either very few or very many practice trials are used.

Overall, then, these studies of instructional manipulation of subjective organization with unrelated words have shown quite mixed results. Instructing Ss to recall alphabetically produces a facilitation of recall, but there is good reason to believe that this is due as much to the cue function of first letters as to the systematic ordering of recall that is produced. Instructing

Ss to use any of a variety of bases for developing a fixed order of recall (a high degree of subjective organization) has shown that they can significantly increase the level of such organization, but has not resulted in the expected augmentation of recall. These findings thus do not strongly support the view that organization determines the amount that can be remembered. On the ther hand, the investigations along the lines started by Mandler do suggest that instructing Ss to categorize words is a sufficient condition for the retention of those materials.

There have also been a few instructional manipulation studies using lists of conceptually categorized, or related, words. Only a single known study has followed the same basic procedure used with the unrelated materials; namely, that of instructing the Ss to organize their recalls according to categories built into the list by E. Newman (1967) utilized this kind of manipulation with lists of CCC trigrams. The lists comprised three categories of three trigrams, where the categories were defined on the basis of common first letters shared by three trigrams. Subjects were given either standard instructions deemphasizing the importance of order or instructions which either described the categorical composition of the list, or told Ss to cluster items from the same category in recall in addition to giving the category information. It was found that the total number of items recalled correctly over all trials was not influenced by the type of instruction, but Ss in the group told to cluster took fewer trials to reach one errorless repetition of the list than did Ss in the other two groups. In terms of the amount of clustering actually observed in recall, the group told to organize their recalls in clusters was again superior, but in this case the group given the categorical information was somewhat inferior to the standard group.

It is unfortunate that there is only this single study of this particular kind. There is a concern about the extent to which findings with these stimulus materials can be readily generalizable to the situation where more highly meaningful materials are used. More specifically, the stimulus materials here comprised atypically short lists, CCC trigrams rather than words, and categories defined on the basis of formal features of the items rather than pre-established verbal relationships. Furthermore, the results of the study are somewhat ambiguous. The group told to cluster did, in fact, show greater organization, but contradictory conclusions about the augmentation of the amount recalled are reached by examination of items recalled as opposed to trials to one errorless repetition.

While the evidence which emerges from the studies with unrelated words and that with related words is far from conclusive, there does not appear to be any reason to doubt the inherent value of the instructional manipulation paradigm.

And because this paradigm has been applied with categorized materials only in the single study by Newman, the purpose of the present experiment was to repeat essentially the same basic design, but with more meaningful stimulus materials.

Method

Two 40-word stimulus-lists were used in this study. One list comprised four taxonomic categories of ten words each (4C/10W) and the other comprised ten categories of four words (10C/4W). The members of each category were drawn from the lower frequency levels of the Battig and Montague (1969) cultural norms and the two lists were perfectly matched in this respect. The lists were also closely balanced on the basis of the mean number of letters per word as well as the mean Thorndike-Lorge (1944) frequency of occurrence.

The stimulus lists were presented by means of a booklet technique. The first page of the booklet requested information concerning age, sex, and any previous experience in experiments. The next page contained the instructions. All subjects received instructions to the effect that they were to study a list of words and that they would then have to write as many of the words as they could remember. The three major experimental groups were distinguished by the nature of the details of the instructions given about the composition of the list and the way in which the items were to be remembered. One group was given standard instructions for free recall. That is, they were given no information about the categorical composition of the list; they were told simply that they would have no difficulty in recognizing any of the words since they would all be quite common and would already by very familiar. They were further instructed that the order in which they recalled the words was not important and that they could write them in any order that was easy for them. This group is designated NI-NC, indicating that they were given no information about categories or about clustering during recall. A second group, designated I-NC, was told that they would notice that the words were taken from several different kinds of categories of familiar things, but they were further instructed that the order in which they recalled the words was not important. The third group, I-C, was given the information about the categorized nature of the list, and was further instructed to try very hard to cluster together the words from the same category, but not to worry if they were not able to do this perfectly.

Subsequent pages in the booklets contained, alternately, the words for study and blanks for the recall of the words. Word presentation pages showed the words listed in the center in a single column. The words appeared in a different randomized order on each trial. One minute periods were allowed for the study of the words. Word recall pages contained a column of 40 blank lines down the center of the page. Periods of one minute were also allowed for recall. All subjects were instructed to write one word per line and to work down the column without skipping any lines as they went. A series of five such study-test trials was administered.

The subjects were tested at the start of regular class periods. A systematically alternating sequence was used to distribute the booklets for the three types of instructions. Six separate groups were tested, with the group sizes varying from 10 to 22 subjects.

The subjects were summer school students at Millersville State College. A total of 93 subjects was tested, but eight had to be omitted from the analysis for failure to follow instructions. Seven subjects had to be dropped from the I-C group, one from the NI-NC group, and none from the I-NC group. The number of subjects included in the results for the NI-NC, I-NC, and I-C groups was respectively 15, 14, and 11 for the 4C/10W list, and 16, 17, and 12 for the 10C/4W list.

Results

The mean number of words recalled on each trial by each of the six experimental groups is shown in Table 16. Means for the three instructional conditions, collapsed over the two types of lists, are also shown. These data were analyzed according to an analysis of variance design comprising instructions and type of list as between-subjects variables, and trials as a within subjects variable. The effect of instructions was significant, F (2,79) = 5.80, $p<.05$, as was the increase in recall scores over trials, F (4,316) = 185.28, $p<.001$. Neither the effect of the different lists nor any of the interactions reached significance at the .05 level of confidence. The results of post hoc comparisons by the Scheffé method supported what is apparent from Table 16, namely, that groups NI-NC and I-NC did not differ from each other at the .05 confidence level, but that both of these groups recalled significantly more words than did the I-C group.

An analysis of the number of categories recalled was performed upon the data from the 10C/4W list. The 4C/10W list was judged not to provide sufficient opportunity for variation in scores to permit a meaningful analysis. A category was scored as being recalled, or represented in recall, if one or more instances of that category appeared in the subject's recall protocol. Mean scores for this measure are presented in Table 17. The analysis in this case included instructions as a between-subjects variable and trials as a within-subjects variable. It was found that instructions had a significant effect, F (2,42) = 4.03, p = .025, as did trials,

Table 16.

Mean Number of Words Recalled on Each Trial in Experiment 6

		Trials				
Stimulus List	Instructional Condition	1	2	3	4	5
	NI-NC	10.00	14.20	15.40	17.67	19.13
4C/10W	I-NC	11.43	14.43	17.07	19.14	20.00
	I-C	9.91	12.36	13.64	15.27	17.18
	NI-NC	11.13	14.94	15.56	17.25	18.44
10C/4W	I-NC	9.47	12.12	14.65	15.88	17.82
	I-C	9.33	10.50	11.92	14.67	16.08
	NI-NC	10.58	14.58	15.48	17.45	18.77
Combined	I-NC	10.32	13.16	15.74	17.35	18.81
	I-C	9.61	11.39	12.74	14.96	16.61

Table 17.

Mean Number of Categories Represented in Recall of
the Ten Category List Under All Instructions
in Experiment 6

Instructional Condition	Trials 1	2	3	4	5
NI-NC	6.89	8.50	8.25	8.12	8.12
I-NC	6.47	6.94	7.65	7.65	7.94
I-C	6.33	6.50	6.83	7.17	7.67

F (4,168) = 9.29, $p<.001$, but the interaction of these variables was not significant, F (8,168) = 1.47, $p>.05$. Results of post hoc comparisons revealed once again that the NI-NC and I-NC groups did not differ from each other. However, only the NI-NC group had significantly (.05) more categories represented in recall than observed with the I-C group (i.e., the I-NC and I-C groups did not differ).

Further analyses were designed to examine the extent to which categorical clustering was present in recall under the different conditions. Every recall sequence was examined to determine the number of observed category repetitions, O(SCR). A unit of O(SCR) was scored each time a word from any category was directly followed in recall by another word from the same category.

It is clear that the value of O(SCR) is not independent of other parameters of the recall sequence from which it arises, such as the total number of words recalled, the number of categories represented in recall, and the number of words recalled within each of the categories represented in recall. Recently, a number of different types of derived scores have been developed in an attempt to control for variations in these other parameters, making possible thereby an unconfounded comparison of the degree of clustering found in different recall sequences. Two different types of derived scores were used in this study.

The first type of score was the observed-minus-expected SCR difference-score, O-E(SCR), as described by Bousfield and Bousfield (1966). According to their formula, the number of units of SCR which are expected on the basis of chance, E(SCR), is a function of the number of words recalled in each of the categories.and the total number of words recalled. The O-E(SCR) difference-score was found for every recall sequence of every subject. Mean scores for this measure are shown in Table 18. These data were subjected to the same three-way analysis of variance. The only significant effects revealed by the analysis were those for lists, F (1,79) = 13.17, $p<.01$, and trials, F (4,316) = 66.92, $p<.001$. Thus, the observed clustering deviated from the expected values significantly more in the case of the 4C/10W list than with the 10C/4W list, and in general the scores increased as a function of trials. The effect of instructions was not significant at the .05 level (though it did approach the necessary value) and no reliable interactions were obtained.

The second type of derived score which was calculated here was the so called adjusted ratio of clustering (ARC) as described by Roenker, Thompson and Brown (1971). The ARC score simply expresses the actual O-E(SCR) deviation relative to the maximum possible O-E deviation (i.e., ARC=O-E/Max-E).

Table 18.

Mean O-E(SCR) Clustering Scores for Each Trial in Experiment 6

Stimulus List	Instructional Condition	Trials 1	2	3	4	5
	NI-NC	1.72	2.97	5.03	6.34	7.98
4C/10W	I-NC	2.05	3.42	5.10	6.33	8.12
	I-C	1.04	2.21	2.68	3.95	5.28
	NI-NC	.47	1.52	2.04	3.40	5.14
10C/4W	I-NC	.34	1.99	2.76	3.63	5.13
	I-C	.63	1.37	1.34	2.23	3.51

Mean values of the ARC score are shown in Table 19. The analysis of these data revealed only a significant increase in scores over trials, F (4,316) = 15.47, $p<.001$. No other sources of variation achieved significance at the .05 level of confidence.

The final clustering measure used in this study was the raw number of clusters observed in each recall protocol. The number of clusters is defined by the number of runs of items regardless of their length, or alternatively, as the number of times the subject switched categories. It must be noted that this measure is not linearly related to the magnitude of the clustering phenomenon as it is intuitively understood or as it is defined in terms of SCR units. However, it was used in this study to compliment the other measures in an attempt to get a fuller specification of the performance in this situation. Once again the analysis was restricted to the 10C/4W list, and the mean number of clusters for each of the instructional conditions is shown in Table 20. These data were subjected to the analysis of variance with instructions as the only between-subjects variable and trials as a within-subjects variable. This analysis indicated that instructions had a significant effect, F (2,42) = 3.73, $p<.05$. The effect of trials was also significant, F (4,168) = 13.21, $p<.001$, but the instructions x trials interaction was not significant, F (8,168) = 1.84, $p>.05$. In order to isolate the instructional effect post hoc comparisons by the Scheffé method were performed. These comparisons revealed that the recall sequences of the I-C group comprised significantly fewer clusters than was the case for the NI-NC group, and that no other comparisons were significant at the .05 confidence level.

Discussion

The first results which must be considered are those for the O-E(SCR) and ARC clustering scores. The non-significant effect of instructions upon both of these measures reveals that instructions were not effective in inducing any differences in the degree of clustering imposed upon recall. Furthermore, the absence of significant interaction terms indicates that instructions were equally ineffective at manipulating clustering in both the 10C/4W list and the 4C/10W list, as well as across all five trials. Thus, even when subjects were given instructions to try to cluster their recalled items they produced no greater deviation of observed levels of clustering from the expected amounts either in terms of absolute deviation scores (O-E) or relative (ARC) deviation measures. In short, the attempted instructional manipulation of clustering was apparently not effective.

On the other hand, the results for the amount of material recalled did show significant variation due to instructions.

Table 19.

Mean ARC Clustering Scores for Each Trial in Experiment 6

Stimulus List	Instructional Condition	1	2	3	4	5
4C/10W	NI-NC	.405	.355	.594	.576	.722
	I-NC	.260	.378	.496	.548	.635
	I-C	.326	.368	.401	.511	.601
10C/4W	NI-NC	.126	.276	.277	.419	.609
	I-NC	.265	.513	.471	.552	.601
	I-C	.245	.287	.224	.239	.435

Table 20.

Mean Number of Clusters Observed in Recall of the Ten Category List Under All Conditions in Experiment 6

Instructional Condition	Trials 1	2	3	4	5
NI-NC	9.81	12.38	12.25	12.31	11.69
I-NC	8.47	9.18	10.65	10.82	11.18
I-C	8.00	8.33	9.50	11.00	11.08

It was found that the I-C group recalled significantly fewer words than either the NI-NC or the I-NC groups. Once again, these findings held equally well with both lists and across trials. It was concluded, then, that the subjects who were given instructions to cluster their recalled items actually recalled fewer words than subjects in the other groups.

This combination of results is, at least initially, perhaps somewhat surprising. All known previous attempts to manipulate the degree of organization of recall through variations in instructions were successful regardless of whether they involved unrelated lists (Mayhew, 1967; Puff, 1970; and Tulving, 1962b) of categorized lists (Newman, 1967). Furthermore, the present results represent an apparent contradiction of the very popular notion that the amount that can be recalled is critically dependent upon the degree of organization (e.g., Tulving, 1962a; Mandler, 1967). And, in this respect, there is some basic agreement with most of the previous research using this same basic experimental paradigm. Here, groups which did not differ in the degree of organization manifest in recall did show differences in the amount of material recalled. In the case of the studies by Mayhew (1967), Newman (1967), and Puff (1970) the instructional groups differed in organization, but not in the amount of material recalled. Thus, the precise pattern of results may vary somewhat but the failure to support the postulated critical role of organization is consistent.

What appears to be the most plausible interpretation of the present results is as follows. The semantically based categories used here were presumably very obvious and potent, thus presenting little problem of detection by the subjects. There is a marked contrast in this respect with the study by Newman (1967) where the stimulus items were CCC trigrams and the categories were defined on the basis of common first letters (i.e., his categories comprised formal rather than semantic relationships). Perhaps it is the case that when given such meaningful semantic categories all Ss normally utilize them and cluster to their maximum capability without the necessity of any help provided by the instructions. A straightforward comparison between the present results and Newman's is virtually impossible for a number of reasons (e.g., Newman didn't report any clustering data for trials before No. 5, individual trial recall and clustering data are not reported, etc.). However, a very liberal estimation of what his typical ARC score for trials 5-16 might have been given a value of about .33, or somewhat less than half of the overall average value for trial 5 in the present study. This is at least consistent with the notion that a relatively high degree of clustering was going on in the present study.

The second element of the explanation is the proposition that the instructions to cluster may have induced those subjects to spend a greater amount of time trying to fill-out a current cluster as completely as possible, and verifying that it included as many words as they could possibly remember at that time, before proceeding to a cluster from another category. During the extra time they spent in these activities the traces of other, as yet unrecalled, items may have faded to too great an extent to permit their recall. The findings of the analyses of the 10C/4W list data showing that the I-C subjects quite consistently had the fewest categories represented in recall and the smallest number of raw clusters (though only the NI-NC group was significantly superior at the .05 level) are at least consonant with this interpretation. Such a tendency would, of course, not be expected to be revealed by the O-E(SCR) and ARC measures if they do their intended job of correcting for differences in the various parameters of amount of recall.

In sum, the proposed explanation is that the instructions to cluster were effective in modifying the subjects' behavior in that it induced them to invest more time in each cluster they produced. Since these particular stimulus materials presumably already elicited a maximal clustering tendency from all subjects regardless of instructions, the I-C subjects, in effect, wasted time which could have been used to recall additional words.

This explanation was clearly derived from the precise pattern of observed results, and is uncomfortably circular. However, it does suggest a number of kinds of implications which are quite testable in further investigations. Perhaps the most obvious possibility would be to keep a record of the temporal course of recall by individual subjects. This would indicate whether the recall by the I-C subjects showed longer pauses supportive of the postulated greater amount of time spent by these subjects at the end of one cluster and before the beginning of the next. A second interesting type of study would be to essentially repeat the conditions of the present investigation with the addition of a group of subjects who were given the information about the categories, but who were instructed to cluster as little as possible. This should avoid some of the problems encountered in the present study where, for example, it is possible that the degree of clustering for all subjects could not be exceeded by the I-C subjects. Observing recall performance when a significant decrease in clustering has been induced should provide just as adequate a test of the "organization hypothesis" as was afforded by the conditions of the present study. Finally, if the explanation of the results obtained here is correct, the more expected pattern of results might be obtained by repeating the same experiment with stimulus materials designed to elicit less clustering. That is, less potent categories could be used leaving thereby more of an opportunity to augment the level of clustering when the instructions to that effect are administered.

Incidentally, it might be noted that the present results for the two types of lists corroborate Weist's (1970) general findings. The relative clustering score which he described, and which has subsequently been called the ARC score by Roenker, Thompson and Brown (1971), is indeed more independent of the parameters of the recall sequence than is the simple O-E score.

Experiment 7: Effects of Categorical and Non-Categorical Context Upon Recall of the Same Stimulus Words

Purpose

It is commonly observed that the presentation of a list of conceptually categorized material (C list) results in a greater proportion of the material being recalled than is the case with a stimulus list which does not comprise such categorical relationships (NC list). Results like these have been observed in studies by Fisher (1971), Puff (1970), and Roberts (1968), to cite just a few of the possible illustrations. It is also typically proposed that the advantage of the C list is due to the fact that the categorical structure of this kind of list provides a readily detectable and highly effective basis for organizing the material, forming strategies for its retrieval from memory storage, and/or for cueing the recall of items which might otherwise have been forgotten. In other words, it has been emphasized that the special advantage of the C list accrues from the subjects' use of the previously learned relationships (associations, or verbal habits) among the items of the same category.

In interpreting findings like these it is important to draw a distinction between two kinds of properties of verbal stimuli. One such property is a relationship between two or more items. This is, most generally, referred to as association strength. Association strength is usually assessed and specified through the use of a normative technique involving some kind of association procedure such as free association or restricted association, where the S must respond with a particular type of response. The restricted association procedure most relevant to the present kind of study is one where the response is to be a member of the class, concept, or category, represented by the stimulus item (e.g., see Battig & Montague, 1969; Cohen, Bousfield & Whitmarsh, 1957). A second type of characteristic of verbal stimulus material pertains only to individual items. This property is variously referred to as meaningfulness, familiarity, vividness, imagery value, etc. It is typically assessed and defined through the use of norms based upon frequency of occurrence of the individual words in samples of the language (e.g., Thorndike & Lorge, 1944), the number of other words emitted as associations to the stimulus word (e.g., Noble, 1952), the rated imagery value (Paivio, Yuille & Madigan, 1968), or the rated vividness of the word (Tulving, McNulty & Ozier, 1965).

These normative properties of individual items have also been found to influence the amount of stimulus material that can be remembered. For example, amount of recall has been seen to vary as a function of meaningfulness (McGeoch, 1930), frequency of occurrence (Hall, 1954), vividness (Tulving, McNulty

& Ozier, 1965), and imagery value (Paivio, Yuille & Rogers, 1969). Results like these clearly support the necessity of taking individual item properties into account when attempting to account for the amount recalled in any situation.

Previous comparisons of C and NC lists have all had a common basic design. Separate C and NC lists are prepared. The C list is constructed so that it has a higher degree of item interrelationships, or association strength, according to an appropriate set of norms. In fact, in order to be strictly objective, the categories in the C list exist only to the extent indicated by the norms. Additionally, one or more properties of the individual items are statistically controlled or held constant so that the mean values for the C and NC lists are the same. Any observed differences in amount recalled are subsequently attributed to the differences in item interrelationships.

Much of the work in the area of verbal learning and memory rests upon the adequacy of norms for manipulating the experimental task. There is certainly a great deal of evidence to support the use of norms for predicting with a fair degree of accuracy what performance will obtain with a given set of materials (e.g., see Bousfield, Steward & Cowan, 1964).

On the other hand, it is also clear that the use of norms has some important weaknesses and limitations. Perhaps the most striking kind of illustration of this is the observation that two lists which are equated on the supposedly relevant normative properties lead to quite different behavior during the recall task (e.g., see Puff, 1972). Postman (1963) has pointed out another kind of limitation, namely, that it is virtually impossible to create a list of words which are unrelated in the eyes of Ss in spite of what is indicated by the norms. Furthermore, Cofer (1967) has shown that in several cases where investigators have prepared lists comprising associations between category names and category instances these lists have also, quite unintentionally, embodied extensive associations among the category instances themselves. The basic implication of findings like these would appear to be that when a very important conclusion is largely based upon normative manipulations, it ought to be verified in other, more direct, ways as well.

Thus, the purpose of the present study was to provide a somewhat clearer test of the role of the two kinds of properties of verbal items in the recall advantage of C lists. Here, the frequency of recall of exactly the same words was investigated as a function of whether they were presented in the context of items to which they were related (C list) or in the relative absence of related items (NC list). In this way, some of the inherent dangers in statistical control of individual item properties were avoided.

Method

A practice list of 18 numbers was randomly chosen from those between 0 and 50. These numbers were written as words in a single randomized sequence for presentation to Ss. The experimental materials consisted of two C and two NC lists. The C lists both comprised 10 taxonomic categories (sports, vehicles, animals, etc.) of 4 members each. These items were drawn from the associations given to the category names in the Battig and Montague (1969) norms. The mean cultural frequencies for the categories in the two C lists are shown in Table 21. Additionally, an attempt was made to keep all lists as homogeneous as possible in terms of Thorndike-Lorge (1944) frequencies of occurrence and mean number of letters per word. Means of these values for all four lists are also shown in Table 21.

The lists were also constructed so that some words appeared in both the C and NC lists. More specifically, one word from each category in one of the C lists also appeared in one of the NC lists. That is, 10 words from the C1 list were also in the NC1 list, and 10 words from the C2 list also appeared in the NC2 list. These words are referred to here as common items. The remaining 30 words in each list were unique to that list and are referred to as filler items. Five separate randomized orders of presentation of each of these lists were used.

The data were collected by means of a booklet technique. Separate booklets were prepared for each of the four experimental lists. Testing was done in four intact class groups. Booklets of each type were haphazardly distributed in each class. Because of the nature of some of the anticipated analyses, more C list than NC list booklets were administered.

Instructions for the practice list were studied by the subjects as they were read aloud by the experimenter. These instructions were to the effect that the Ss were to study the column of items and that they would subsequently be asked to write as many of the items as they could remember. It was emphasized that the order in which they wrote the items was not important and that they could write them in any order which seemed easy or "natural" for them. The practice list items appeared in a single column in the center of the next page. This list was studied for one minute before the subjects turned to the next page and wrote all they could remember in a period of one minute. Then the instructions for the experimental list were presented. These were identical to those for the practice list, but stressed that the practice list bore no relationship to the list that they were about to see. A study period of two minutes was allowed for the experimental list, and two minutes were given for writing the recalled words on the next page.

Table 21.

Some Characteristics of the Stimulus Lists
and the Number of Subjects Presented With Each List
in Experiment 7

	Lists			
	C1 List	NC1 List	C2 List	NC2 List
Mean Category Association Frequency (Battig & Montague, 1969)	20.6		18.1	
Mean Frequency of Occurrence (Thorndike-Lorge, 1944)	27.4	29.7	30.2	30.5
Mean Number of Letters per Word	5.5	5.6	5.9	5.4
Number of Ss for List Condition	30	21	30	20

The subjects were 101 undergraduate students at Millersville State College. The number of booklets completed for each of the four list conditions is shown in Table 21.

Results

The amount recalled from the *practice list* was analyzed first. Mean recall scores for the *Ss* who subsequently received each of the four different experimental lists are shown in Table 22. The analysis of variance design involved type of list (C vs. NC) and list-pair no. (1 vs. 2) as between *Ss* variables. The results of the analysis revealed no significant variation due to type of list, F (1,97) = <1.0, list-pair, F (1,97) = 1.18, $p>.05$, or the interaction of the two variables, F (1,97) = 1.35, $p>.05$. It was therefore concluded that the four groups of subjects were quite homogeneous with respect to their basic ability for this type of task.

The first treatment of the experimental list data involved the total number of words recalled. Mean total recall scores for each of the experimental groups are shown in Table 22. These data were then subjected to the same 2 x 2 design as outlined above. As expected on the basis of previous findings, significantly more words were recalled from the C than from the NC lists, F (1,97) = 5.72, $p<.025$. No significant variation in these scores was attributed to the specific list-pairs, F (1,49) = 0.00. Finally, while the C1 list resulted in an average of about 3.6 more total words recalled than the NC1 list, and the C2 list exceeded the NC2 list by only about .5 words, the interaction between type of list and list-pair was only marginally significant, F (1,97) = 3.35, $.10>p>.05$. There was no evidence of significant heterogeneity of variance or of any linear relationship between means and variances so transformation of the data was not in order. However, there is a sizeable question about whether lists which are matched only to the extent that they are in the present study should be included as a dimension in the analysis of variance. Since there is no definitive way to answer this question, and since the list-type x list-pairs interaction was marginally significant, separate comparisons between C and NC lists were carried out for each pair of lists. The results of these comparisons supported the conclusion that significantly more words were recalled from C1 than from NC1, t (49) = 3.33, $p<.01$, while there were no differences in recall from C2 and NC2, t (48) = .36, $p>.20$.

The mean recall scores for *filler words only* (i.e., those 30 words in each list which were not common to another list) are also presented in Table 22. The treatment of these data showed exactly the same pattern of results as was found for total list recall. According to the analysis of variance, once again more filler words were recalled if they were from C lists than if they were from NC lists, F (1,97) = 4.46, $p<.05$. Additionally, neither the effect of list-pair

Table 22.

Mean Recall Scores of Various Kinds of Items From
Categorized and Non-Categorized Lists
in Experiment 7

Type of Item	Stimulus List			
	C1	NC1	C2	NC2
No. of Practice Items	10.40	9.67	10.37	10.65
Total No. of Items	16.30	12.67	14.73	14.25
No. of Filler Items	11.77	9.14	11.00	10.70
No. of Common Items	4.43	3.52	3.73	3.55
Prop. of Filler Items	.39	.30	.37	.36
Prop. of Common Items	.44	.35	.37	.35

(1 vs. 2), nor the interaction of list-pair x type of list, F (1,97) = 2.82, .10>p>.05, reached the .05 level of confidence. As before, the separate comparisons revealed that significantly more filler words were recalled from C1 than from NC1, t (49) = 3.20, p<.01, while C2 and NC2 were quite equivalent in this respect, t (48) = .27, p>.20.

Mean recall scores for common words only (i.e., those 10 words in each list which appeared in both a C and a NC list) are shown in Table 22. Much the same pattern emerged once again. In this case, the superiority of C over NC lists was substantially reduced, so that the main effect of type of list was now only marginally significant, F (1,97) = 3.27, .10>p>.05. The effect of list pairs was not significant, F (1,97) = 1.24, p>.10. The strength of the type of list x list-pairs interaction was also reduced so that it was now not even marginally significant, F (1,97) = 1.45, p>.10. Since the use of separate comparisons was previously justified on the basis that they might be more appropriate even on a priori grounds, they were employed again with these data even though they could not be justified by the magnitude of the type of list x list-pairs interaction observed in the analysis of variance. A significantly greater number of the common words were recalled from C1 than from NC1, t (49) = 2.09, p<.05, but equal numbers of the words common to lists C2 and NC2 were recalled from those lists, t (48) = .44, p>.20.

The final treatment of the data involved comparisons of the recall of common vs. filler items within each of the lists. Since the recall scores for the filler items could have gotten as large as 30 while those for the common items could not exceed 10, observed values of both scores were converted to proportions of the maximum possible values. Mean proportion scores are shown in Table 22. Comparisons within each list by means of t-tests for related means were performed with results as follows: for the C1 list, t (29) = 1.69, p>.10; for the NC1 list, t (20) = 1.51, p>.10; for the C2 list, t (29) = .19, p>.10; and for the NC2 list, t (19) = -.03, p>.10. In short, just about the same proportion of common and filler items were recalled within each list.

Discussion

It was expected that, in keeping with the preponderance of the relevant previous research, more words would be recalled from the C than from the NC lists, and that this would hold as well for the recall of just filler items as for total items recalled. Either of two possible outcomes were entertained for the common items. Firstly, it was expected that these items also might be recalled better after presentation in the C than the NC lists. This kind of outcome is, of course, consistent with the view that it is the interitem relationships among the members of a category which are the basis for the superior recall of categorized materials. In other words, according to this view the provision of other members of a category as context should result in augmented recall of the common items from the C lists. In contrast to this, it was also anticipated that the common items might be recalled equally well from the C and NC lists. This outcome would support the view that the reason that greater recall is typically observed with C lists is because the items selected for these lists are individually more memorable than items in the NC lists, even though efforts are usually made to match the lists on some normative basis. In this case then, it is the properties of the individual items which are critical while the categorized or non-categorized nature of the context within which the items are presented is irrelevant.

Unfortunately, the observed results do not conform to either expectation definitively, and the nature of the conclusions depends upon several statistical issues. Thus, if it is assumed that the pairs of C and NC lists constitute a meaningful dimension in the analysis of variance, and if the conventional .05 level of confidence is strictly adhered to, one set of conclusions is reached. That is, these results indicate that significantly more total words and more filler words were recalled from C than NC lists, but that common items were recalled equally well from C and NC lists. Furthermore, there were no significant overall effects of pairs of lists, nor any significant interactions indicating any differential C-NC effects in the two list pairs. Looked at this way, the results thus seem pretty straightforwardly to deny the importance of related context and to support the role of individual item properties.

If the two list pairs are treated entirely separately (i.e., if they are regarded essentially as replications rather than levels of a variable) and the comparisons are made by $\underline{t}$-tests a quite different pattern of results obtains. In this case, when the results for the C1 and NC1 lists are considered it is found that significantly more total, filler, and common words are recalled from the C1 list. However, when the C2 and NC2 lists are considered it is seen that these lists were quite equivalent in the recall of total, filler, and common words. The separate comparison of C1 and NC1 thus gives clear support for the notion that interitem relationships and related context are important to the superiority of recall of categorized materials. Since for some unexplainable reason the expected overall superiority of the C list was not found in the C2 vs. NC2 comparison, the assessment of the number of common words recalled from these lists is probably not very meaningful for the present purposes.

Finally, if marginally significant ($.10>\underline{p}>.05$) results in the analysis of variance are considered, the pattern of results is more similar to that derived from the separate comparison approach. That is, the marginally significant interactions in the analysis of total words and filler words suggests that the C list superiority is greater in the case of C1 vs. NC1 than for C2 vs. NC2. However, then we find that, for common words, the main effect of C vs. NC is marginally significant while the interaction is not. This suggests greater recall of common items from the C lists and that this effect holds equally well for both list pairs. This, of course, contrasts with the results of the separate comparisons.

General Discussion and Conclusions

The Role of Organization in Memory

Probaly the most crucial issue which the studies in this project bear upon is the question of the nature of the role of organization in free recall. As indicated in the Introduction, the most widely accepted current view is that the memory for an amount of material which exceeds the immediate memory capacity of approximately 7±2 individual words is critically dependent upon the effective use of organizational processes or strategies by the subjects (e.g., Mandler, 1967, 1968; Tulving, 1962a, 1968). In the simplest terms, it is thought that as organization goes, so goes the amount of recall. There is certainly a large amount of evidence which is consistent with this view, but there are also a number of empirical contradictions (e.g., Allen, 1968) and alternative theoretical interpretations (e.g., Cofer, 1967). Thus, six of the present experiments continued the assessment of this theoretical view from a number of different approaches.

One class of evidence which has traditionally been interpreted as supporting the view that recall performance is facilitated by organization is the well replicated finding that more words are recalled from a list comprising common conceptual categories (C list) than from a non-categorized list (NC list) whose members do not belong to obvious normative categories (Cofer, 1967; Wood, 1968; Wood & Underwood, 1967; etc.). The basis for inferring support of the "organizational theory" is that conceptual clustering is commonly observed in the recall of the C list, and the augmented recall of these materials could be attributable to the occurrence of this type of organization.

The results of Experiment 1, however, failed to support this interpretation. It was found that when the subjects given the C list were divided into those who showed significant clustering and those whose clustering did not exceed a chance amount, both of these groups recalled significantly more words than did the subjects given a NC list. Furthermore, there were no significant differences in the amount recalled from the C list by the "clusterers" and the "non-clusterers". These results thus do not provide any evidence that the amount that can be recalled is dependent upon the degree of clustering imposed in recall.

Findings like these are, however, consistent with interpretations by Cofer, Wood, and Underwood, that the superiority of C list recall must involve some mechanism other than clustering on the basis of category names at the time of output. Cofer has suggested that the augmented recall with the C list reflects an advantage in terms of interitem associative relationships among the stimulus items. Wood and Underwood have proposed that the superior recall with the C list is attributable to the occurrence of common implicit associative responses which result in an increased availability, via backward associations, of the representational responses for the individual stimulus words. This "priming" which facilitates the recall of the words, is assumed to occur during the study of the material and prior to the time of recall. The results of Experiment 7, which tended to support the importance of interitem associations over the properties of individual words, can also be viewed as generally consistent with these interpretations.

Several possible limitations of the results of Experiment 1 need to be emphasized. First of all, it is possible that the conclusions to be drawn from this study are limited to organization in the form of clustering--the gross grouping of items according to broad, experimenter-defined, conceptual categories. Secondly, the discrepant results in the recent study by Thompson, Hamlin, and Roenker (1972) need to be kept in mind. The investigation of the large number of procedural variations between their study and the present one should ultimately indicate one or more additional important limits to the generality of the results of both experiments.

The work on Experiment 6 represents a second type of approach to testing for the dependence of recall performance upon organization. In this case, an effort was made to experimentally induce different degrees of organization in the form of category clustering through the use of different sets of instructions and then to observe the consequent influence of this manipulation upon the amount recalled. The primary advantage of this approach is that all other conditions of the experimental situation (e.g., stimulus materials, time intervals, etc.) remain constant for all subjects.

The results of this study, unfortunately, revealed that the instructional manipulation of the degree of clustering was not effective. Subjects given instructions to cluster did not give evidence of having used any more of this form of organization than did those not explicitly instructed to employ clustering. Interestingly, however, it was found that the subjects given instructions to cluster actually recalled significantly fewer words than the subjects not given such instructions. In short, groups which did not differ in the amount of organization did differ significantly in the amount of recall.

These results clearly do not fit very well with the general "organizational theory" position. It is possible to generate an alternative explanation in an admittedly post hoc fashion. Briefly, the proposed explanation is that the instructions to cluster were effective in modifying the subjects' behavior in that it induced them to invest more time attempting to fully complete each cluster they produced. Since the very potent categories in the stimulus materials perhaps already elicited a maximal clustering tendency from all subjects regardless of instructions, the subjects given the instructions to cluster might have, in effect, wasted some time which could otherwise have been used for the retrieval of additional words before they were forgotten.

These findings both contrast and agree with those of some previous studies. Mayhew (1967) and Puff (1970) manipulated subjective, or intertrial, organization while Newman (1967) manipulated clustering with lists of 9 CCC trigrams with the categories defined on the basis of the first letters. Instructional groups were found to differ in organization in all of these studies, but to be equivalent in amount of recall. Thus, the precise pattern of results may vary somewhat, but the failure to support the postulated critical role of organization is consistent.

Results consistent with the view that recall performance is dependent upon organization in the form of conceptual clustering were obtained in Experiment 4. This study compared the effects of several different types of orders of presenting categorized materials. It was found that both in terms of the degree of clustering and the amount of material recalled, a blocked sequence where all members of a category appeared in direct succession, was superior to both a random arrangement, and a systematically alternating sequence (i.e., one with properties like ABCABC).

In addition to providing general support for the "organizational theory" these results also bear upon the nature of the more specific kinds of mechanisms which have been proposed to account for the effects of different types of input organization. The obtained pattern of results seems to indicate that the important dimension of input organization is one of proximity, or contiguity, between conceptually related items, while the overall degree of sequential structure, as it was specified here anyway, appears to be a considerably less important dimension. There are a number of different theoretical mechanisms for why proximity is important, and the results of this experiment offer no apparent way to distinguish between them. The supported kinds of views range from notions that implicit mediating responses (Puff, 1966), or implicit representational responses (Wood & Underwood, 1967), are made more available, to speculations that greater proximity increases the probability that related words will be in short-term storage together (Glanzer, 1969), or results in a greater amount of effective processing time (D'Agostino, 1969). Since these results indicate that overall sequential list structure is not an important dimension of input organization, they do not seem to support views which stress the utilization of list structure as a basis for coding material for storage or for the development of plans for the retrieval of information from storage in the manner proposed by Bower, Clark, Lesgold & Winzenz (1969), Cohen (1970), Lewis (1971), and Newman (1967).

Whereas the previous experiments studied organization in the form of categorical clustering, Experiment 5 dealt with subjective, or intertrial, organization. The time between recall of one word and the recall of the next was found to be significantly shorter when the words comprised a unit of subjective organization than when the words were not defined to be organizationally linked. Thus, words defined as a single unit on the basis of the usual defining operation of repeated contiguous recall were also found to show close temporal contiguity as well. These results were interpreted as supporting the "reality" of measured units of subjective organization. In other words, these results seem to indicate some further evidence that the measurement of subjective organization does tap at least some aspects of the functional structure of the subject's memory.

Several other aspects of these results lend additional support to the general view of the importance of organization in memory. For one thing, the distinction in times between words in organizational units and between words which were defined as not organizationally linked only appeared after the subjects had recalled a number of words which was approximately equal to the 7±2 item immediate memory capacity. This is in keeping with the theory that the important function of organization is in allowing the amount retained to exceed the individual item capacity of this short-term storage system. Furthermore, the decrease in the very long interword times seen toward the end of the recall sequences in the early practice trials seemed to be attributable to the increasing density of organizational units during later practice trials. However, it is somewhat disturbing that there was no evidence of organizational units having become stronger or more integrated, with repeated usage; the number of them simply became greater.

Finally, it should be mentioned that in Experiment 3 no differences were observed in the amount of material recalled after either a standard serial type of presentation and study or a simultaneous method. It is at least minimally consistent with "organizational theory" that there were also no differences observed in either clustering or subjective organization.

Overall, then, the studies in this report which are relevant to the issue of whether the amount of material that can be recalled is crucially dependent upon the degree of organization imposed by the subjects show mixed results.

The findings of Experiments 1, 6, and to some extent, 7, run counter to this view. On the other hand, the findings of Experiments 3 and 4 are consistent with "organizational theory", and Experiment 5 provides rather strong support for this view. None of these studies, and none yet performed by any other investigators, constitutes a critical test of this theory. Thus, the process of evaluating this theory must be one of continuing to accumulate evidence until the weight of that evidence points convincingly toward acceptance or rejection of this view. For the present, the existence of negative evidence like that obtained here indicates, at the very least, the necessity for further basic research, and a great deal of caution about advancing a general proposition that recall is dependent upon organization.

It is perhaps instructive to speculate how the evidence can continue to come out both positive and negative. The most optimistic possibility, of course, is that the theory is basically correct and the available measurement techniques are adequate, but that there are some important boundary conditions which delimit the domain of applicability of the theory. That is, perhaps the theory, as stated, does not apply under certain experimental conditions. As a very simple illustration, perhaps some stimulus materials are already so organized along dimensions which we do not yet understand that additional organization by the subject is not required.

Another possibility is that "organizational theory" is correct, but that the available techniques for specifying organization do not tap the "real" nature of the structure of organization in memory. In other words, the critical organization presumably occurs prior to the time that the items are produced in the output phase of a recall trial, and we attempt to infer the nature of this organization by examination of the ordering of recall. There may well be characteristics of the output process which tend to obscure the true nature of the implicit organization so that it is never accurately reflected in the recall sequence, or it may be that we have not yet discovered an adequate index of recall ordering. This state of affairs could explain why the least impressive performances of "organizational theory" in accounting for data have occurred in situations where organization has actually been measured by some clustering or subjective organization index, rather than being inferred from the results of, for example, some transfer operations. It should be noted, however, that results like those found in Experiment 5 do tend to argue against this kind of an interpretation somewhat. Still, this possibility suggests the need for continued attempts to examine new techniques for assessing organization in recall. Some potentially promising developments along these lines could be the use of Johnson's (1967) hierarchical clustering scheme analysis and Allen's (1971) application of the graph theoretic approach.

A final interesting possibility, suggested by Tulving (1968), is that recall performance and organization are not causally related, but are both independent manifestations, or reflections, of some third factor. The nature of this "higher-order" factor cannot yet be specified, but presumably it would be something very roughly analogous to Jensen's (1971) associative or conceptual abilities. Since recall performance and organization might both be manifestations of this more basic factor it would be expected that they would normally be correlated, but there might well be instances where they are not equivalently elicited or revealed in performance. This kind of a view might therefore be better able to incorporate some of the existing evidence which is so embarrassing to the current version of "organizational theory". This view is certainly worthy of further consideration.

Comparisons of Two Types of Organization

A second major contribution of several studies on this project concerns the comparison of organization in the form of categorical clustering and the repeated sequential ordering called subjective, or intertrial, organization. The results of Experiment 2 provide the most direct kind of comparative evidence. It was found that clustering in recall of C lists occurred to a significantly greater extent than did intertrial organiaation in recall of either C or NC lists. Furthermore, the level of intertrial organization with the C and NC lists was quite equivalent.

These results thus seem to suggest that clustering on the basis of conceptual categories is adopted much more readily than is the development of a relatively fixed order of recall even though the subjects have the opportunity to use either or both forms of organization. These results were interpreted as indicating that clustering is in some sense simpler or easier. That is, clustering requires the utilization of relatively obvious, broad, conceptual categories to encompass all the items. In contrast, the best available evidence about subjective organization is that most subjects tend to utilize a number of different techniques for forming units (Abramczyk & Bousfield, 1967; Bousfield & Abramczyk, 1966). Various units may be formed on the basis of story devices, rhymes, input list contiguity, associative relationships, purely idiosyncratic associations, etc. Thus, the imposition of equal degrees of the two forms of organization would presumably require much more organizational, or mediational, activity for the typical subject in the case of subjective organization. It is consistent with this interpretation that those subjects who seem to use only a single basis for subjective organization, such as alphabetization or an elaborate story device, are usually successful in producing a level of organization quite equivalent to that of clustering.

The second implication of the major findings of Experiment 2 is that the two forms of organization appear to be at least somewhat independent of each other. The occurrence of a very high level of clustering in the recall of the C list did not reduce the degree of subjective organization with the same list below that found in recall of the NC list. Thus, contrary to Tulving's (1962a) speculation, clustering did not appear to attenuate the simultaneous use of subjective organization. However, intuition and some work by Mandler (1969) indicated that there are some limitations to the independent operation of the two forms of organization. For example, if the subject is somehow induced to use a serial ordering strategy by a technique such as giving him only one new item on each trial, and if the order of presenting new items comprises a randomized sequence of categories, then the serial ordering will preclude the occurrence of clustering. If the order of presenting the new items is systematically arranged with the members of the same category following each other, then the serial ordering should augment the clustering. Furthermore, it should be found that the clustering of categorized pairs of words will elevate the level of subjective organization over that found here. Also, if the stimulus materials involve only a few very large categories it might well be expected that serial ordering within clusters would be more important. In short, while the two forms of organization were seen to operate quite independently here it is possible to anticipate a number of instances where it can be expected that they could interact in various ways.

The comparison of the results of Experiment 5 for units of subjective, or intertrial, organization with the findings for clustering by other investigators provides some further evidence of the similarities and differences between the two forms of organization. One of the main findings of Experiment 5 was that words in ITR units were emitted in closer temporal contiguity than words not in ITR units, but only after the subject had recalled a number of words approximately equal to the postulated immediate memory capacity. In other words, the effects of subjective organization became progressively greater in the later stages of the recall sequence. Virtually the same pattern of results was found for organization in the form of category clustering by Pollio, Richards and Lucas (1969). Thus, the temporal characteristics of the two forms of organization appear to be quite similar.

The distribution, or density, of ITR units across the recall sequence was also investigated in Experiment 5. It was found that the density of these units of subjective organization decreased gradually, but consistently, across successive stages of the recall sequence, and while practice markedly increased the overall density of such units it did not change the nature of the way they were distributed across stages of recall. A rough comparison can be made between these findings and data for clustering provided by Bousfield and Cohen (1953) though any differences in results might be attributable to any of a number of procedural differences between the studies as well as to inherent distinctions between the two types of organization. First of all, in keeping with the results of Experiment 2, the overall level of clustering in the Bousfield and Cohen experiment was higher than that observed in the present study of intertrial organization. Furthermore, practice did modify the nature of the distribution of units of clustering across stages of recall. More specifically, practice served primarily to increase the density of clustering in the early stages of the recall sequance, so that after a few trials the density of clustering decraased across the stages of recall in a quite negatively accelerated form. Thus, there is at least some similarity here in the occurrence of units of organization in that both decrease monotonically across stages of the recall sequence after a few practice trials.

Experiment 3 also involved both clustering and subjective organization. In this case it was found that neither type of organization was differentially affected when the stimulus items were presented one at a time for serial study or all at once for simultaneous study by the subjects.

Experiment 6 indicated that instructions explicitly telling subjects to cluster words from the same category were not effective in raising the observed level of clustering. These results contrast with several studies which have shown instructions to be quite successful in manipulating the level of subjective organization seen in recall (Mayhew, 1967; Puff, 1970; Tulving, 1962b). However, these results probably do not deserve a great deal of emphasis because of the possibility that the very obvious and potent categories in Experiment 6 may have elicited the maximum possible clustering effort from all subjects regardless of instructions. The fact that Newman (1967) did find effects of instructions upon clustering strengthens this possibility since his stimulus materials involved categories defined on the basis of common first letters of CCC trigrams instead of meaningful semantic relationships.

In sum, it would appear that clustering and subjective organization are basically similar in that they show close temporal contiguity within units and are both distributed monotonically across stages of the recall sequence. Also, they were both similarly non-differentially affected by serial vs. simultaneous presentation and study conditions.

However, there is still some question about the extent to which the two can be influenced by instructions since the precise nature of the categorized stimulus materials may be an interacting factor. The only really striking difference between them is that clustering appears to be utilized to a much greater extent than is subjective organization under the conditions investigated here. Furthermore, under the present conditions, the two forms of organization appeared to be used substantially independently of each other.

Comparisons of Measures of Organization

The number of available measures, or indices, of organization is expanding rapidly. This is particularly true in the case of clustering measures, but a similar trend appears to be beginning in the specification of subjective organization. As indicated before, the major problem is that the amount of organization observed in a particular recall protocol depends to some extent upon other features of that protocol such as the total number of words recalled, etc. What is needed, in order to make meaningful comparisons between subjects or conditions, is a measure which takes into account the possible contribution of chance, and/or what the range of possible observed values could be in each case. In other words, the values of the measure of organization should be independent of the other characteristics of the recall sequences.

The earliest and most widely used measures, the Ratio of Repetition (Bousfield, 1953) and the observed-minus-expected deviation scores (Bousfield & Bousfield, 1966) have been heavily criticized on these kinds of statistical grounds, and each of the critics has proposed some new kind of measure. The approaches suggested by Dalrymple-Alford (1970), Dunn (1969), Frankel and Cole (1971), Hudson and Dunn (1969), and Roenker, Thompson and Brown (1971) have already been reviewed and are summarized in Appendix B. These measures all seem to have both advantages and disadvantages which are revealed by analytical arguments, artificial experiments, and a few demonstration experiments. None of these efforts have yet provided convincing evidence that one type of measure is substantially superior or preferable to the others. Consequently, different investigators use different measures.

The controversy over the adequacy of the various measures raises two serious questions. One has to do with the confidence that can be placed in the body of information about organizational processes that was accumulated with the earlier measures. The second question concerns the confidence with which results of current studies using the various measures can be integrated and compared.

The ultimate answers to questions like these will, of course, require some definitive analytical work. However, an empirical approach can also provide some relevant evidence and that is the procedure followed in the present project. Wherever feasible, the results were analyzed with more than one measure to provide some evidence about the effects of the different measures and to increase the range of comparability of the findings of the study.

The broadest investigation of this sort was carried out in Experiment 4 where the effects of different orders of presenting categorized materials were assessed with seven different measures of clustering. These measures included both of the older measures and all five of the newer measures mentioned above. An identical pattern of results was found in every case. No significant differences were found between randomized and systematically alternating types of input orders, and both of these led to significantly less clustering than did a blocked presentation of the categories.

The similar pattern of results with all of the different clustering measures supports the findings of earlier studies which used the older measures and found that the blocked presentation resulted in greater clustering than did a randomized presentation. Thus, the findings of the earlier studies of this particular phenomenon cannot be considered to represent some kind of statistical artifact arising from the nature of the measures used at that time. This also lends some confidence to the findings of earlier clustering studies in general, though it is far from certain that the same findings would hold in the case of other independent variables. These findings indicate that, under these circumstances at least, the choice of measures is not very critical since they do not reveal any special advantage for any of the measures.

Only two of the most popular measures were employed in Experiment 6. This study involved the attempted instructional manipulatin of clustering using lists of four categories of ten words and lists of ten categories of four words. The results were analyzed with the observed-minus-expected deviation scores of Bousfield and Bousfield (1966) and the adjusted ratio of clustering (ARC) score of Roenker, Thompson, and Brown (1971) which was previously called the relative clustering score by Weist (1970).

In this case, the two measures did show different patterns of results. The use of the traditional deviation scores indicated that the clustering with the list of four categories of ten words was greater than with the list of ten categories of four words, whereas the ARC scores indicated no significant difference between the two types of lists. The amount recalled with the list of four categories of ten words was somewhat, though not significantly, greater and the two lists differed (by design) in both the number of categories and the number of words per category represented in recall. The clustering results thus seem to imply that the deviation scores were influenced by these parameters while the ARC scores were much more independent of them. These findings therefore confirm Weist's (1970) work. They also indicate that the choice of measures is quite important in this kind of situation. Furthermore, according to the criterion of independence of recall parameters, the ARC measure would appear to be preferable to the more common deviation score.

It was also shown in Experiment 2 that the choice of measures is quite critical in cases where the two forms of organization are to be compared as directly as possible. The observed-minus-expected deviation scores were to be basically inappropriate for this type of comparison because of the fact that this scale has different end points for the two types of organization. A score expressing observed organization relative to the maximum possible was found to be adequate, but probably still does not represent the best possible procedure.

Overall, then, the comparison of various measures has indicated that the choice of measures is not very critical in some cases, as in the study of input order effects, but is important when the number of categories and items per category vary between conditions, and is most critical of all when the two forms of organization are to be directly compared.

Conditions Affecting the Use of Organization

Experiment 3 revealed that neither clustering nor subjective organization were differentially affected by serial and simultaneous methods of presenting the material. It was expected that presenting all of the material for simultaneous study would result in greater organization because the subjects would have a greater opportunity to detect the structure of the material, to scan back and forth looking for related items, etc. Before greatly emphasizing the fact that the two presentation and study methods are functionally equivalent the possibility that the task was too easy should be evaluated further.

The use of clustering was found, in Experiment 4, to vary with the order of presentation of the categorized materials. Specifically, orders of presentation in which the members of the same category were blocked together led to more clustering than did either systematically alternating or randomized orders of presentation.

The degree of clustering was not found to vary in Experiment 6 as a function of the instructions given to the subjects. Equivalent degrees of clustering were imposed in recall by groups explicitly instructed to try to cluster, those given information about the categorical structure of the list, and those given no information about the structure of the list or clustering. Once again, however, a really firm conclusion is precluded by the possibility that the categories used in this study were so potent as to elicit the maximal clustering effort regardless of instructions.

Thus, of the conditions investigated, the order of presenting categorical material did affect the use of organization in the form of clustering. Instructions were not successful in manipulating clustering, and neither form of organization was affected by the type of presentation and study method.

The Booklet Technique

The data for Experiments 1, 4, 6, and 7 were collected by means of a booklet technique. That is, each subject was given a booklet containing the instructions, the material to be learned, and blanks for recording the recalled responses. Where more than a single practice trial was used, successive pages in the booklet alternately presented the material again and provided blanks for the recall tests. This kind of experimental booklet is thus roughly equivalent to a workbook, and was administered in regular classroom settings.

It was reasoned that the booklet technique would make the experimental task more analogous to familiar educational procedures and would therefore be more "natural" for the subjects. This, in turn, should increase the potential generalizability of the findings of this kind of basic research to the more applied situation. The booklet technique is also a very efficient means of data collection. On the other hand, it cannot be used where the experiment requires exceedingly precise control of exposure duration, etc. Most studies of human learning and memory do not require an exceptional degree of control over such factors, and the booklet has much to recommend it's more general usage.

References

Abramczyk, R. R. & Bousfield, W. A. Sequential ordering in free recall as a function of interitem associative strength. Psychological Record, 1967, 117, 183-192.

Allen, M. Rehearsal strategies and response cueing as determinants of organization in free recall. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 58-63.

Allen, M. Graph theory, organization, and memory. Paper presented at meetings of the Eastern Psychological Association, New York, N. Y., April, 1971.

Battig, W. F. & Montague, W. E. A replication and extension of the Connecticut category norms. Journal of Experimental Psychology Monograph, 1969, 80, 1-46.

Bousfield, W. A. The occurrence of clustering in the recall of randomly arranged associates. Journal of General Psychology, 1953, 49, 229-240.

Bousfield, W. A. & Abramczyk, R. R. Sequential ordering in repeated free recall as a function of the length of the stimulus word list. Canadian Journal of Psychology, 1966, 20, 427-434.

Bousfield, A. K. & Bousfield, W. A. Measurement of clustering and of sequential constancies in repeated free recall. Psychological Reports, 1966, 19, 935-942.

Bousfield, W. A. & Cohen, B. H. The effects of reinforcement on the occurrence of clustering in the recall of randomly arranged associates. Journal of Psychology, 1953, 36, 67-81.

Bousfield, W. A., Puff, C. R., & Cowan, T. M. The development of constancies in sequential organization during repeated free recall. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 489-495.

Bousfield, W. A., Steward, J. R. & Cowan, T. M. The use of free associational norms for the prediction of clustering. Journal of General Psychology, 1964, 70, 205-214.

Bower, G. H., Clark, M. C., Lesgold, A. M. & Winzenz, D. Hierarchical retrieval schemes in recall of categorized word lists. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 323-343.

Cofer, C. N. Does conceptual organization influence the amount retained in immediate free recall? In B. Kleinmumtz (Ed.), Concepts and the structure of memory. New York: John Wiley & Sons, 1967, pp. 181-214.

Cohen, B.H. Long term retention of categorized word lists. Paper presented at the meetings of the Eastern Psychological Association, Atlantic City, April, 1970.

Cohen, B. H., Bousfield, W. A., & Whitmarsh, G. A. Cultural norms for verbal items in 43 categories. Technical Report No. 22, 1957, University of Connecticut, Contract Nonr-631(00), Office of Naval Research.

Cole, M., Frankel, F. & Sharp, D. Development of free recall learning in children. Developmental Psychology, 1971, 4, 109-123.

D'Agostino, P. R. The blocked-random effect in recall and recognition. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 815-820.

Dalrymple-Alford, E. C. The measurement of clustering in free recall. Psychological Bulletin, 1970, 1, 32-34.

Dunn, J. E. A compounded multiple runs distribution. Journal of the American Statistical Association, 1969, 64, 1415-1423.

Earhard, M. The facilitation of memorization by alphabetic instructions. Canadian Journal of Psychology, 1967, 21, 15-24.

Elmes, D. G. & Wright, J. H. Effects of verbal recall of blocking trigrams of similar connotative meaning. Journal of Experimental Psychology, 1970, 85, 11-15.

Fisher, D. F. The effects of delay interval on word recall and clustering. Journal of Psychology, 1971, 77, 67-77.

Foote, R. A. & Pollio, H. R. The role of associative organization in free recall under five different presentation conditions. Psychonomic Science, 1970, 20, 205-207.

Frankel, F. & Cole, M. Measures of category clustering in free recall. Psychological Bulletin, 1971, 76, 39-44.

Glanzer, M. Distance between related words in free recall: Trace of the STS. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 105-111.

Glanzer, M. & Clark, W. H. Accuracy of perceptual recall: An analysis of organization. Journal of Verbal Learning and Verbal Behavior, 1962, 1, 289-299.

Hall, J. F. Learning as a function of word frequency. American Journal of Psychology, 1954, 67, 138-140.

Heimer, W. I. & Tatz, S. J. A comparison between massed and distributed simultaneous display and a serial procedure in free-recall learning. Paper presented at the meetings of the Eastern Psychological Association, Atlantic City, April, 1970.

Hudson, R. L. & Dunn, J. E. A major modification of the Bousfield (1966) measure of category clustering. Behavioral Research Methods and Instrumentation, 1969, 1, 110-111.

Hultsch, D. F. Adult age differences in the organization of free recall. Developmental Psychology, 1969, 1, 673-678.

Jensen, A. R. The role of verbal mediation in mental development. Journal of Genetic Psychology, 1971, 118, 39-70.

Johnson, S. C. Hierarchical clustering schemes. Psychometrika, 1967, 32, 241-254.

Lewis, M. Q. Categorized lists and cued recall. Journal of Experimental Psychology, 1971, 87, 129-131.

Mandler, G. Organization and memory. In Spence, K. W. & Spence, J. T. (Eds.), The Psychology of Learning and Motivation. New York: Academic Press, 1967, 1, 327-372.

Mandler, G. Association and organization: facts, fancies, and theories. In T. R. Dixon and D. L. Horton (Eds.), Verbal Behavior and General Behavior Theory. New Jersey: Prentice-Hall, Inc., 1968, 109-119.

Mandler, G. Input variables and output strategies in free recall of categorized lists. The American Journal of Psychology, 1969, 82, 531-539.

Mandler, G. Words, lists, and categories: An experimental view of organized memory. In J. L. Cowan (Ed.), Studies in Thought and Language. Tucson, Arizona: The University of Arizona Press, 1970, pp.99-131.

Mayhew, A. J. Interlist changes in subjective organization during free-recall learning. Journal of Experimental Psychology, 1967, 74, 425-430.

Mayzner, M. S. & Tresselt, M. E. Incidental learning: A function of associative strength and distance between S-R pairs. Journal of Psychology, 1961, 53, 155-160.

McGeoch, J. A. The influence of associative value upon the difficulty of non-sense syllable lists. Journal of Genetic Psychology, 1930, 37, 421-426.

McLean, R. S. & Gregg, L. W. Effects of induced chunking on temporal aspects of serial recitation. Journal of Experimental Psychology, 1967, 74, 455-459.

Miller, G. A. Information and memory. Scientific American. 1956, 195, 42-46. (a)

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, 63, 81-97. (b)

Murdock, B. B., Jr., & Okada, R. Interresponse times in single-trial free recill. *Journal of Experimental Psychology*, 1970, 86, 263-267.

Nelson, K. J., McRae, B. C. & Sturges, P. T. Recall as a function of instructions and trials. *Journal of Experimental Psychology*, 1971, 90, 152-153.

Newman, S. E. Multitrial free recall: Effects of clustered presentation, element-sharing, and instructions. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 844-850.

Noble, C. E. An analysis of meaning. *Psychological Review*, 1952, 59, 421-430.

Oldfield, R. A. Memory mechanisms and the theory of schemata. *British Journal of Psychology*, 1954, 45, 14-23.

Paivio, A., Yuille, J. C. & Madigan, S. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 1968, 76 (1, Pt. 2).

Paivio, A., Yuille, J. C. & Rogers, T. B. Noun imagery and meaningfulness in free and serial recall. *Journal of Experimental Psychology*, 1969, 79, 509-514.

Payne, B. The relationship between a measure of organization for visual patterns and their judged complexity. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 338-343.

Pellegrino, J. W. A general measure of organizaiton in free recall for variable unit size and internal sequential consistency. *Behavior Research Methods and Instrumentation*, 1971, 3, 241-246.

Pollio, G. R., Richards, S., & Lucas, R. Temporal properties of category recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 529-536.

Postman, L. Does interference theory predict too much forgetting? *Journal of Verbal Learning and Verbal Behavior*, 1963, 2, 40-48.

Puff, C. R. The measurement of clustering. Paper presented at the Eastern Psychological Association meetings, New York, April, 1963.

Puff, C. R. Clustering as a function of the sequential organization of stimulus word lists. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 503-506.

Puff, C. R. An investigation of two forms of organization of two forms of organization in free recall. Journal of Verbal Learning and Verbal Behavior. 1970, 9, 720-724.

Puff, C. R. Temporal properties of organization in recall of unrelated words. Journal of Experimental Psychology, 1972, 92, 225-231.

Puff, C. R. & Bousfield, W. A. Development of intertrial organization with simultaneous presentation of stimulus items. Journal of Verbal Learning and Behavior, 1966, 5, 503-506.

Quaintance, B. & Shapiro, S. I. Class membership and clustering in free-recall learning. Canadian Journal of Psychology, 1970, 24, 70-76.

Roberts, W. A. An analysis of multitrial free recall learning with input order held constant. Journal of Psychology, 1968, 68, 22--242.

Roenker, D. L., Thompson, C. P. & Brown, S. C. Comparison of measures for the estimation of clustering in free recall. Psychological Bulletin, 1971, 76, 45-48.

Rosner, S. R. The effects of rehearsal and chunking instructions upon children's multitrial free recall. Journal of Experimental Child Psychology, 1971, 11, 93-105.

Shapiro, S. I. & Bell, J. A. The formation of organizational units in free recall learning. Psychonomic Science, 1971, 22, 127-218.

Shuell, T. J. Clustering and organization in free recall. Psychological Bulletin, 1969, 72, 353-374.

Slamecka, N. J. Testing for associative storage in multitrial free recall. Journal of Experimental Psychology, 1969, 81, 557-560.

Spitz, H. H. The role of imput organization in the learning and memory of mental retardates. In Ellis, N. R. (Ed.), International Review of Research in Mental Retardation. New York: Academic Press, 1966, pp. 29-56.

Sturges, D. T., Crawford, C. B. & Nelson, A. K. The effect of instructions upon free recall and subjective organization of abstract and concrete words. Paper presented at the Midwestern Psychological Association, Detroit, May, 1971.

Thompson, C. P., Hamlin, V. J., & Roenker, D. L. A comment on the role of clustering in free recall. Journal of Experimental Psychology, 1972, 94, 108-109.

Thorndike, E. L. & Lorge, I. *The teacher's word book of 30,000 words* New York: Teachers College, Columbia University, 1944.

Tulving, E. Subjective organization in free recall of "unrelated" words. *Psychological Review*, 1962, 69, 344-354 (a).

Tulving, E. The effect of alphabetical organization on memorizing unrelated words. *Canadian Journal of Psychology*, 1962, 16, 185-191 (b).

Tulving, E. Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal Behavior and General Behavior Theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Tulving, E., McNulty, J. A. & Ozier, M. Vividness of words and learning to learn in free-recall learning. *Canadian Journal of Psychology*, 1965, 19, 242-252.

Tulving, E. & Pearlstone, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 381-391.

Underwood, B. J. & Freund, J. S. Further studies on conceptual similarity in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 30-35.

Wallace, W. P. Consistency of emission order in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 58-68.

Weingartner, H. The free recall of associatively related words. *Journal of Verbal Learning and Verbal Behavior*, 1964, 3, 6-10.

Weist, R. M. Optimal versus nonoptimal conditions for retrieval. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 311-316.

Winograd, E., Conn, C. P., & Rand, J. Superiority of complete presentation to single-item presentation in recall of sequentially organized material. *Journal of Experimental Psychology*, 1971, 88, 223-230.

Wood, G. Implicit responses and conseptual similarity: a repetition. *Journal of Verbal Learning and Verbal Behavior*, 1968, 7, 838-840.

Wood, G. & Underwood, B. J. Implicit responses and conceptual similarity. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 1-10.

Yoshimura, E. K., Moely, B. E. & Shapiro, S. I. The influence of age and presentation order upon children's free recall and learning to learn. *Psychonomic Science*, 1971, 23, 261-263.

Appendix A.

Listing of Experiments Performed, Relationship to Original Proposal, and Dissemination Activities

Experiments as Numbered and Titled in this Report	Most Closely Corresponding Study in Original Proposal	Dissemination Activities
1. Role of clustering in free recall.	2. Recall of identical items from categorized and non-categorized word lists.	Submitted as Interim Report No. 1. Published in the Journal of Experimental Psychology, 1970, 86, 384-384.
2. An investigation of two forms of organization in free recall.	1. Clustering and intertrial organization in recall of the same material.	Submitted as Interim Report No. 2. Published in the Journal of Verbal Learning and Verbal Behavior, 1970, 9, 720-724. Presented at meetings of Eastern Psychological Association, Atlantic City, April, 1970.
3. Free recall with serial and simultaneous presentation of categorized word-lists.	4. Clustering in recall after simultaneous presentation of stimuli.	Submitted as Interim Report No. 3.
4. An investigation of memory performance with three types of presentation sequences and seven measures of organization.	3. Clustering as a function of list length.	Submitted as Interim Reports Nos. 4 and 5.* Presented at meetings of Eastern Psychological Association, Boston, April, 1972.

Appendix A. (cont'd)

Experiments as Numbered and Titled in this Report	Most Closely Corresponding Study in Original Proposal	Dissemination Activities
5. Temporal properties of organization in recall of unrelated words.	5. The temporal relationship between consistent recall and entry into units of organization.	Submitted as Interim Report No. 6. Published in the Journal of Experimental Psychology, 1972, 92, 225-231. Presented at meetings of Eastern Psychological Association, New York, April, 1971.
6. Effects of instructions upon clustering with conceptually categorized materials.	X. No close correspondence to any specific experiment.	Submitted as Interim Report No.7.
7. Effects of categorical and non-categorical context upon recall of the same stimulus words.	2. Recall of identical items from categorized and non-categorized word lists.	Submitted as Interim Report No. 8.

---

* Interim Report No. 5 represents a lengthy review of the literature relevant to the experiment covered in Report No. 4.

Appendix B.

Brief Formulas for the Clustering Measures

Common Terminology

O(SCR) = Number of observed stimulus category repetitions.
E(SCR) = Number of repetitions expected by chance.
Max(SCR) = Maximum possible SCR value.
Min(SCR) = Minimum possible SCR value.
N = Total number of words recalled.

Measure	Investigator(s)	Formula
Ratio of Repetition	Bousfield (1953)	$RR = \frac{O(SCR)}{N-1}$
O-E(SCR) Difference	Bousfield & Bousfield (1966)	$O-E(SCR) = O(SCR) - E(SCR)$
Adjusted Ratio of Clustering	Roenker, Thompson, & Brown (1971)	$ARC = \frac{O(SCR) - E(SCR)}{Max(SCR) - E(SCR)}$
D_A Index	Dalrymple-Alford (1970)	$D_A = \frac{O(SCR) - E(SCR)}{Max(SCR) - Min(SCR)}$
Standardized SCR Score	Hudson & Dunn (1969)	$Z(SCR) = \frac{O(SCR) - E(SCR)}{Var(SCR)}$
Z-Score Runs Test	Frankel & Cole (1971)	$Z(Runs) = \frac{O(Runs) - E(Runs)}{Var(Runs)}$
D_N Index	Dunn (1969)	$D_N = \frac{O(SCR) - E(SCR)^{*}}{Var(SCR)}$

* While this measure and Hudson & Dunn's look identical in this summary form, they involve differences in the calculation of the E(SCR) and Var(SCR) terms.